UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ARTIFICIAL INTELLIGENCE(AI)-INTEGRATED COMPUTER-AIDED DIAGNOSIS(CAD) SCHEMES FOR DEVELOPING PROGNOSTIC MODELS IN STROKE STUDIES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ROWZAT FAIZ

Norman, Oklahoma

2024

ARTIFICIAL INTELLIGENCE(AI)-INTEGRATED COMPUTER-AIDED DIAGNOSIS(CAD) SCHEMES FOR DEVELOPING PROGNOSTIC MODELS IN STROKE STUDIES

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. David Ebert, Chair

Dr. Qinggong Tang

Dr. Samuel Cheng

Dr. Javier Jo

Dr. Zahed Siddique

To my Father, Dr. Md. Faiz Ullah

"Abbuji, as I wrote this dissertation, all I wished for
was to hear your words of support and prayer.
More than anything, I wish I could share this
accomplishment with you."

To my Mother, Nazma Faiz, for your endless love and strength.

To my Husband, Dr. Shajid Islam, my rock and inspiration.

Acknowledgement

I would like to start by expressing my heartfelt gratitude to my advisor, Dr. David Ebert, for his unwavering support and advice during this journey. I am especially thankful to him for his crucial support during a pivotal phase of my PhD, which was instrumental in helping me reach this milestone. I am also profoundly grateful to Dr. Gopichandh Danala, whose invaluable insights have been fundamental to the completion of this work.

My heartfelt thanks to Dr. Zahed Siddique for not only serving on my committee but also for providing steadfast support during a critical time in my PhD journey. It has been a true privilege to have Dr. Qinggong Tang, Dr. Javier Jo, and Dr. Samuel Cheng on my PhD committee. Their insightful guidance and encouraging feedback was vital to the completion of my doctoral work. I also want to thank Dr. Bin Zheng, a former committee member, for his kind guidance and support before his departure from the university.

I would like to extend my special gratitude to the Assistant Director of the ECE Department, Rodney Keele, who has always welcomed me with a warm smile, listened patiently to any concerns, and provided the most generous support possible. It has also been an honor to be a part of the Data Institute for Societal Challenges (DISC-OU) team, and I am grateful for the friendships and experiences I have shared with the people there. I will always cherish the memories we created together during group meetings, discussions, and special occasions.

My heartfelt thanks go to Proma Chowdhury, Tasfiq E Alam, and Warid Islam—friends who have become like family to me here in Norman, OK. This path has been challenging, but the moments we shared made the journey not only easier but truly unforgettable. I am so grateful for your presence in my life. Forever thankful to my dearest friend Maria Hossain for her positive energy and encouragement. A special thanks to my best friend Keya Das, whose enduring moral support has been a constant source of strength.

Finally, the core of my strength, my family. I am forever indebted to my late father, Dr. Md. Faiz Ullah, and my mother, Nazma Faiz, for believing in me and taking a leap of faith to support my journey by all means. Coming from a small town in Bangladesh, achieving a dream like this would have been impossible without the unwavering support of parents like

them. I have no words to express my sincere gratitude for their incredible sacrifices.

I am thankful to my brother(-in-law), Dr. Masud Chowdhury, my god-send guardian angel, and my sister, Rizwana F. Chowdhury for everything they both have done for me and for being my roof in the U.S.A. To my nephew and niece, Istaeen and Durdana, thank you for your unconditional love. I am deeply obliged to my mother-in-law, Rebeka Sultana, whose encouragement and prayers have inspired me to reach this milestone. I also fondly remember my late father-in-law, Rafiqul Islam, who would have been one of the proudest today.

Last but certainly not least, I am profoundly grateful to my loving husband, Dr. Shajid Islam, for his unyielding support through every step of this journey. Even during the most challenging times, his faith in me and his constant encouragement kept me grounded and moving forward. I am forever indebted to him for the sacrifices he made and the burdens he carried alone along the way.

Above all, I offer my deepest gratitude to the Almighty, who deserves the highest reverence for all that I have been blessed with.

Table of Contents

List of Figures

List of Tables

Abstract

The field of computer-aided diagnosis and detection (CAD) systems is rapidly expanding, with the integration of artificial intelligence (AI) significantly enhancing its capabilities. The development assists the doctors in detecting potential abnormalities, diagnosing the condition, and/or possible prognosis by examining the abnormal region. The CAD systems hold the potential to predict clinical events and future outcomes by employing biomarkers extracted from clinical data and/or imaging data. In this dissertation, the studies primarily focused on predicting the prognostics of multiple stroke conditions employing quantitative imaging biomarkers. The objective was to develop robust AI integrated CAD, AI-CAD, systems that would leverage advanced techniques to improve stroke outcome prediction accuracy and prognostic efficiency, specifically for Subarachnoid Hemorrhage (aSAH) stroke and Acute Ischemic Stroke (AIS) patients. This predictive capability would assist doctors in planning in-hospital treatment, allocating resources effectively, and guiding individualized follow-up care. Such an approach not only addresses critical needs requiring immediate attention but also offers financial benefits to many patients.

Quantitative radiographic biomarkers predict prognosis in several neurological diseases. However, aSAH with its varied clinical course has limited objective tools to predict clinical outcomes accurately. The first study of the dissertation focused on developing a CAD scheme that explored the efficacy of a deep-learning classification architecture in predicting several short-term clinical events and functional outcomes, utilizing brain-computed tomography (CT) scans for aSAH patients. A retrospective dataset encompassing 60 aSAH patients was collated, each with two CT images acquired upon admission and after 10-14 days of admission (referred to as discharge in the study). The in-house developed CAD scheme was utilized for segmenting and labeling the brain regions and selecting CT slices based on the presence of blood clots in the relevant cisternal spaces. The transfer learning method was applied with two pre-trained architectures, DenseNet-121 and VGG16, selected as convolutional bases for feature extraction. The Convolution Based Attention Module (CBAM) was integrated atop the pre-trained architecture to enhance focus learning. Employing five-fold cross-validation,

the developed prediction model assessed three clinical outcomes following aSAH: two short-term clinical events typically occur within 72 hours of initial onset and one future functional outcome. The performance of the model was evaluated using multiple metrics. A comparison was conducted to analyze the impact of CBAM comparing the performance from the base models and CBAM integrated models. In the study, it was observed that the prediction model trained using CT images acquired at admission demonstrated higher accuracy in predicting short-term clinical outcomes. Conversely, the model was trained using CT images acquired on 10/14 days to accurately predict long-term clinical outcomes. Notably, for short-term outcomes, high sensitivity performances (0.88 and 0.86) were reported from the admission scan, while the sensitivity of (0.53 and 0.61) was reported from the discharge scan. This result showcased the viability of predicting the prognosis of aSAH patients using novel deep learning-based quantitative image markers. The study demonstrated the potential of integrating deep-learning architecture with attention mechanisms to optimize predictive capabilities in identifying clinical complications among patients with aSAH.

To create a more clinically viable prediction model, it was necessary to incorporate domain knowledge alongside automatically extracted hierarchical deep features.. With that objective, the second study of this dissertation focused on analyzing the performance of an ensembled hybrid model with fused features that combine the strengths of radiomics and deep learning techniques to objectively improve aSAH prognosis prediction. In the study, radiomic engineering and deep transfer learning techniques were applied to extract relevant features. Several predictive models, including radiomics, deep transfer learning, and combined ensemble machine learning approaches, were developed to predict four short-term and four long-term clinical outcomes. The analysis also included assessing admission and discharge scans to develop all prognostic models for eight events. Finally, receiver operating characteristic (ROC), accuracy, sensitivity, specificity, F1-score, and precision were computed to assess the model's performance. The results indicated that hybrid ensemble models outperformed individual approaches, indicating complementary information that was extracted improved the prognostic performance. Additionally, the results of this study

also demonstrated that admission scans predict short-term outcomes better, while discharge scans more effectively predict long-term outcomes, which matched the observation from my first study. This study highlighted the potential of quantitative hybrid ensemble CAD as a high-performing prognostic model for evaluating clinical complications in aSAH patients

To broaden the scope of this dissertation, my third study focused on more common types of stroke (AIS) patients. There is limited data on objective radiological parameters to predict functional outcomes in AIS patients. Although the conventional 10-point topographic system Alberta stroke program early CT score (ASPECTS) is widely used, it is limited by inter-rater variability and the difficulty of detecting subtle ischemic changes on pre-intervention scans for clinical decision-making. Post-intervention scans can use these scoring systems to better define final cerebral infarct volume (CIV), an objective radiological marker of functional outcomes in AIS. Thus, this pilot study aimed to develop a composite radiological tool as an imaging biomarker by conducting a quantitative volumetric analysis of ASPECTS/Posterior Circulation (PC)-ASPECTS regions, CIV, and total brain volume (TBV). Initially, ASPECTS and PC-ASPECTS regions were annotated on anonymized scans and mapped to CIV. Various volumetric parameters were extracted, including TBV, CIV, and the ratio of CIV to TBV. These raw features enabled calculations of percentage volumes and facilitated summation and proportion analyses. Premorbid and admission clinical features were integrated into the models. Furthermore, raw and transformed features were used to train multiple classification algorithms. Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), Lasso, and Elastic Net (EN) were applied to identify the most predictive features. Results showed that the Support Vector Machine (SVM) model, using EN-selected transformed features and clinical data, achieved the highest AUC (0.93 ± 0.05). Alternatively, Logistic Regression (LR) and Extreme Gradient Boost (XGB) with RFE-selected transformed features and clinical data performed well (AUC 0.91 ± 0.02). Models trained with transformed features generally outperformed others. Among the tested algorithms, the SVM model with EN-selected features was the top performer. This study highlights the potential of quantitative imaging biomarkers, combining volumetric features with clinical pa-

rameters, in developing a reliable prognostic model for AIS patients, surpassing traditional methods.

This dissertation comprises three studies focused on my primary objective: developing prognostic models for stroke-affected patients. I concentrated on building various prediction models, and optimizing the performance of deep architectures and hybrid ensemble models to predict multiple outcomes. Additionally, I explored diverse feature engineering and data optimization techniques to introduce innovative quantitative imaging markers utilizing machine learning, deep learning, and hybrid systems. The simulations and results from these studies demonstrate the proposed CAD schemes' strong discriminative performance in stroke treatment, offering valuable support to radiologists in disease diagnosis and improving diagnostic accuracy.

CHAPTER 1

Introduction

Over the last few decades, image acquisition technology advancements have significantly improved biomedical image processing algorithms, resulting in large quantities of high-resolution images being accessible at a low cost from various modern medical imaging sources, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Ultrasound [2]. These advancements have aimed to enhance diagnostic accuracy and support clinicians and radiologists in decision-making by gaining a detailed view. However, this has become challenging due to the sheer volume of image data added, as well as the limitations in image interpretation caused by visual noise, search patterns, and reader fatigue, which contribute to inter- and intra-reader variability. To capitalize on the availability of large amounts of digital data, computation-based algorithms known as Computer-Aided Diagnostic (CAD) models were researched and developed for early disease diagnosis and prognostic assessments of a patient's health. This provided radiologists with computer-generated assessments as supplementary "second opinions" to inform their diagnostic decisions [8]. Unlike fully automated computer diagnoses, which were solely based on algorithms, CAD integrates human expertise and domain knowledge with computational ca-

pabilities, enhancing physicians' diagnostic skills by effectively complementing them rather than replacing them.

While numerous performance studies have acknowledged the necessity and significance of CAD in clinical settings, traditional systems have encountered several challenges. These include high development costs, inconsistent clinical evaluations, issues related to workflow and cost-effectiveness, and limitations in specific regions of interest (ROI) [9]. Another significant concern has been the high rate of false positives, resulting in unnecessary biopsies and increased recall rates, which can lead to physical harm from unnecessary screenings and impose a substantial economic burden on the healthcare system. As a result, despite 20 years of commercialization of the system, CAD had hardly achieved widespread acceptance or utilization in clinical practice and continued to face technical challenges in improving detection performance and robustness [10][11]. The performance substantially improved with the evolution of Artificial Intelligence (AI).

1.1 AI-CAD

Earlier, CAD techniques relied on the radiomics method, which involves extracting image features and outcomes, enhancing precision while providing reliable results [9, 12]. The technique involved the computation of manually designed image features from the specific segmented areas and the training of traditional Machine Learning (ML) models to asses specific clinical metrics, such as detecting abnormalities, disease detection, and classification. [13]. In figure 1.1, this era around 1980 is marked as the second AI boom among its three phases[1]. This phase was characterized by the advent of new and exciting technologies, coupled with the promise that computers would soon diminish all persistent issues. However, the manual crafting of features lacks a set of global standards, which has limited the efficient application of these methods in clinical practice[14][15]. As a result, the high expectations were not met, leading to a period of extreme disillusionment often referred to as the"AI winter."

Despite this historical context, at the turn of the 21st century and onwards, radiologists

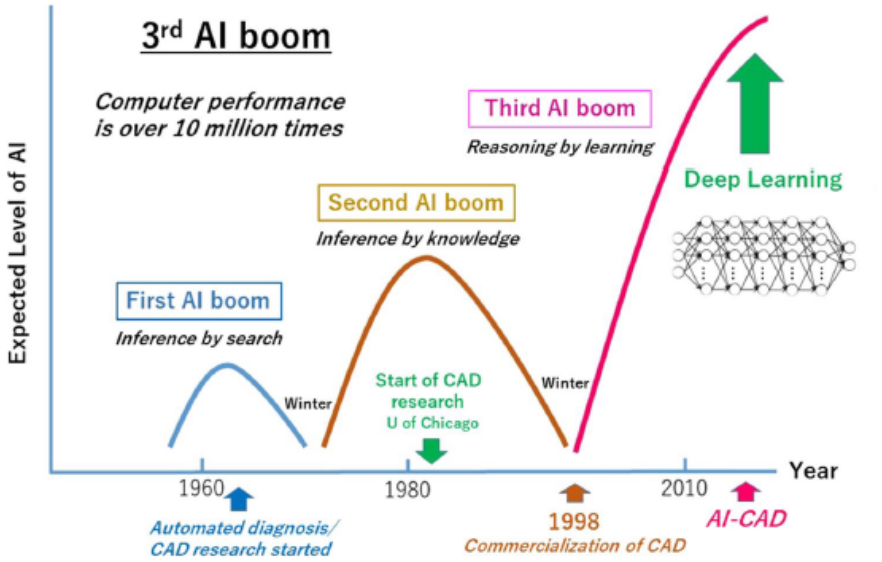

Figure 1.1: Phases of AI boom in CAD development [1]

showed renewed interest in CAD development driven by the development of a new subfield of AI, Deep Learning (DL), that significantly outperformed traditional CAD methods. Modern healthcare has benefited greatly from this advancement in AI, as seen by the development of AI-integrated CAD (AI-CAD) systems that mirror the revolutionary shifts in the diagnostics area that were predicted during the second AI-CAD era. The automated feature extraction ability of DL has made a paradigm shift from the traditional radiomics technique, optimizing the efficiency of workflow, reducing false positives, and improving diagnostic accuracy [16, 17]. The size of the data set was no longer a limitation, owing to the automatic extraction of relevant features as image biomarkers were reported in several modalities [18, 19]. Moreover, besides ML and DL approaches, researchers have recently been exploring hybrid ensemble approaches by combining the strengths of multiple ML algorithms or combining both radiomic and DL techniques to improve prediction accuracy, leading to advancements in different healthcare aspects such as detection, diagnosis, and prediction [20, 21].

1.2 AI-CAD based Prediction Models

Today, AI-driven CAD models have become an integral part of healthcare, improving diagnostic accuracy and assisting in personalized patient care for a range of medical conditions, from cancer to cardiovascular disease [22, 23, 24]. With the development of AI-CAD models for early disease detection, growing interest has been shown in employing CAD beyond the diagnosis and characterization of diseases. These include quantitative analysis of tumor heterogeneity, the identification of cancer subtypes, the staging of cancers, the planning and execution of treatment plans, and predicting the prognosis of treatment outcomes [25][26][27]. In modern healthcare, prediction models are being used in increasing numbers as a tool in clinical decision-making regarding certain health conditions or patient treatment outcomes. These models are aimed at supporting doctors and patients to make informed choices for additional medical attention, including tests, treatments, or changes in lifestyles. The goal is to increase the cost-effectiveness of care and improve individual outcomes by offering customized risk assessments for diagnosis or prognosis [28, 29]. Researchers have studied a variety of emerging ML and DL techniques to propose advanced AI-CAD-based prediction models, which have been widely conducted in various medical fields to predict multiple diseases like heart disease, liver disease, Alzheimer's disease, and various types of cancers for which early detection is crucial for treatment [30]. Among the multiple health conditions where AI-CAD models have shown promise, stroke stands out as a critical area where accurate prediction tools can significantly impact recovery outcomes.

1.3 AI-CAD and Stroke Imaging

Stroke is marked as a major global health concern, accounting for the second-highest mortality rate globally and a major cause of lifelong disability. The report from the World Health Organization indicates that 13.7 million people experience a stroke each year, where 5.5 million of those individuals fall into the fatality rate besides being a primary contributor to the significant and lifelong chronicle disability [31, 32]. Stroke survivors are vulnerable

to functional loss, emotional side effects, and recurrence, all of which affect the quality of life highlighting the need for better stroke management and prevention strategies [33]. Early intervention is an essential key to lowering morbidity and mortality as soon as a stroke occurs [34]. For stroke treatment in hospitals, several clinical assessments such as the Face Arm Speech Test (FAST), the National Institutes of Health Stroke Scale (NIHSS), the Cincinnati Pre-hospital Stroke Scale (CPSS), and neuroimaging methods like MRI and CT scans are traditionally employed. However, accurate diagnosis might be difficult because of missed symptoms or conflicting symptoms with previous strokes [33]. Thus, AI in stroke imaging is a rapidly emerging topic with numerous applications for both hemorrhagic and ischemic stroke subtypes. Numerous studies have reported the development of AI-CAD models for several areas of the stroke treatment model, such as the detection of infarcts or hemorrhages[35], segmentation[36], classification[37], identification of major vascular occlusions[38], grading of the Alberta Stroke Program Early CT Score[39], early prediction of stroke[40] and prognosis of stroke outcomes[41, 42].

Stroke prognostic models have the potential for early and accurate identification of patients at higher risk of poor short-term and long-term outcomes following stroke onset, enabling tailored treatment plans for an individual, improving patient outcomes, and maximizing healthcare resources [43]. Short-term outcomes of stroke patients occur within 72 hours of onset, which is typically caused by rebleeding, vasospasm, and brain edema, causing Early Brain Injury (EBI) while disrupting regular cerebral function. Long-term outcomes impact the physical and cognitive abilities of stroke patients and often affect their emotional stability.

1.3.1 Prognosis model for Hemmorhagic Stroke Patients

Although radiomic imaging biomarkers have presented impressive results in neuroimaging-based prediction studies like Alzheimer prediction [44] and ischemic stroke outcome[45], very few imaging biomarkers have been investigated for predicting both short-term and long-term prognoses of the hemorrhagic stroke patient. Several serum biomarkers were reviewed

for predicting the outcome of subarachnoid hemorrhage (SAH)[46, 47] but they lacked lab standardization [48]. The modified Fisher scale can be used to estimate the volume of blood in ventricles and subarachnoid space but it has limitations of poor interrater agreement [49]. However, the systemic inflammatory response from location-based blood quantification technique shows some promise in predicting the development of EBI and long-term clinical outcomes; nonetheless, this technique also depends on operator skill and semi-automated methods. The potential of DL-based or an ensembled AI-CAD model can be explored to develop an automated prognosis model for SAH patients.

1.3.2 Prognosis model for Ischemic Stroke Patients

Examining ischemic changes in radiological scans obtained after in-hospital treatment is a common method of assessing the functional outcomes of acute ischemic stroke (AIS) [50, 51]. Historically, ischemic stroke severity and recovery have been assessed using 10-point qualitative measures: Alberta Stroke Program Early CT Score (ASPECTS) [52]. Despite being widely used, the traditional 10-point topographic approach has limitations due to inter-rater variability and the difficulty of detecting small ischemia changes [53, 54]. Additionally, multiple studies have suggested that not all ASPECTS regions have equal contributions to the functional outcome [55]. Therefore, a more advanced and systematic strategy for developing a quantitative prognosis model can be explored by integrating volumetric information from clinically relevant regions to address sustaining restriction during the post-assessment of ischemic stroke patients.

1.4 Thesis Statement

The primary goal of this thesis is to develop robust AI-CAD models leveraging advanced AI techniques to improve stroke outcome prediction accuracy, specifically for SAH and AIS patients. These prognostic models play a crucial role in resource optimization, refining treatment approaches, and ensuring tailored follow-up care. Consequently, the primary

motivation behind this thesis was to create a sophisticated model capable of differentiating between positive and negative outcomes with enhanced precision. This thesis addresses the current gaps in automated imaging biomarkers for short- and long-term outcomes in hemorrhagic stroke patients and advances prognostic modeling for ischemic stroke by using quantitative volumetric data.

The thesis makes several key contributions:

- Development of a novel automated imaging biomarker to evaluate SAH patient outcomes.
- Evaluation of a deep transfer learning framework designed to predict four short-term and four long-term outcomes using CT scan data from both the admission and discharge scans. Two pre-trained deep learning architectures were selected for this purpose.
- Examination of the impact of attention mechanisms by incorporating the Convolution Block Attention Module (CBAM) to enhance model performance. Comparison analyses were accomplished between the basic architectures and the attention-integrated architectures to evaluate the effect of attention mechanisms on prognostic model performance.
- Analysis of the impact of different scanning periods on both short-term and long-term outcomes.
- Exploration of a hybrid ensemble approach by integrating traditional radiomics features with deep learning extracted features, creating an enhanced prognostic tool for SAH patients.
- Comparative analysis among radiomics, deep learning, and hybrid ensemble approaches.
- Development of a comprehensive radiological tool combining volumetric information from critical brain regions in AIS patients and their clinical biomarkers to assess functional outcomes.

1.5 Thesis Outline

This dissertation consists of six chapters, including this introduction.

Chapter 2 establishes the foundation for AI-CAD in predictive and prognostic modeling, covering various AI-based models and optimization techniques. It also reviews prediction models for two types of strokes, discussing their background, current practices, and gaps in the literature.

Chapter 3 presents a deep learning-based prediction model using an automated quantitative imaging biomarker for predicting clinical events and outcomes in aneurysmal subarachnoid hemorrhage (a-SAH) patients. This chapter outlines the model's methodology, examines the impact of attention integration on prediction performance, and considers the effects of different scan periods.

Chapter 4 extends the work to an ensemble approach by combining radiomic and deep learning features. It demonstrates how integrating ML and DL features improves the model's accuracy and reliability. Similar to Chapter 3, this chapter also discusses the impact of different scan periods.

Chapter 5 introduces a quantitative prediction model for ischemic stroke outcomes based on the volume of ASPECTS and PC-ASPECTS regions. This chapter explores the volumetric analysis of extracted and transformed features to enhance prognostic prediction in ischemic stroke patients.

Chapter 6 concludes the dissertation by summarizing the newly developed CAD models and discussing potential applications. It also outlines future research directions for continued advancements.

Overall, the dissertation provides a comprehensive examination of stroke prognostic models, combining deep learning, radiomics, and interdisciplinary collaboration to advance stroke assessment, medical imaging, and patient care.

CHAPTER 2

Overview of AI-Based CAD Models for Prediction and Prognosis

2.1 Background

The early study for automated computer diagnosis began in the 1960s as a result of the boom in medical imaging technology, which has provided an enormous amount of medical imaging data for analysis. However, the computerized analysis did not garner much attention, as powerful computers, high-quality digital image data, and computational resources were hard to come by then. Later, in the 1980s, research shifted from automatic computer diagnosis to the development of CAD, which served as a second opinion to assist radiologists in interpreting medical image data[56]. The systematic development of machine learning (ML) and image analysis tools for medical imaging was initiated in the 1980s by Doi et al. at the University of Chicago's Kurt Rossmann Laboratory [8]. Chan et al conducted the first observer study of a computer-aided detection system by detecting clustered microcalcification on mammogram data, which upheld the performance significance of CAD systems as the second eye for any radiologist [57]. In 1998, the U.S. Food and Drug Administration (FDA) approved the first CAD system for mammography, which began the commercializa-

tion of the CAD system with the primary goal of reducing the number of false-negative rates and addressing inter-rater and intra-reader variability. Traditional CAD systems employ machine learning (ML) algorithms, a subset of AI that uses multidisciplinary methods, including statistics and mathematics, guided by domain-specific knowledge to extract relevant handcrafted radiomic features for analysis of both imaging and non-imaging data to detect and classify abnormalities.

However, 92% of CAD applications were mostly used for screening mammography before 2016 and only were commercialized for detecting diseases in some modalities, i.e., chest X-ray, chest CT, and CT colonoscopy. The limitations appeared owing to a high rate of false positive marks, increasing unnecessary examination steps, and analyzing only a specific region of area of an organ [1, 11]. Then deep learning (DL), another subset of AI evaluated from the conventional method, emerged as the state-of-the-art ML method. Deep learning gained significant popularity in late 2012 by winning the ImageNet competition for classification. Although the concept of a machine learning model with image input was proposed by Fukushima in 1980 and the term "deep learning" was coined in 2007 by Hinton et al., it was this competition that marked the breakthrough of the concept, revolutionizing the field of computer vision [58] [59]. By demonstrating spectacular accuracy rates when applied to a dataset of roughly a million images from the web containing 1,000 different classifications, deep learning has become a trending topic among imaging researchers for classification and segmentation purposes.CAD systems integrated and brought a boom in the healthcare industry by recognizing patterns within large quantities of medical data to automatically extract deep features and use domain-specific knowledge to correlate them with the clinical/pre-clinical condition of patients. Today, AI-integrated CAD systems have expanded their role in healthcare beyond detection to include early prediction of diseases and their potential outcomes. This shift has enhanced clinical decision-making speed, improved diagnostic accuracy, and supported tailored treatment planning based on predicted risk.

2.2 Foundations of AI-CAD in Prediction and Prognosis

AI-CAD systems have demonstrated efficacy in aiding radiologists and clinicians in several medical domains by predicting early disease onset and prognosticating clinical outcomes, including recurrence of cardiac arrest or bleeding in the brain, morbidity, death, and intensive care unit admission[60, 61]. Precision medicine, also known as personalized medicine, involves tailoring treatment and preventive measures to the individual needs of each patient. This approach guides disease prevention, diagnosis, and treatment and helps to plan more focused interventions that improve clinical outcomes by classifying patients according to their sensitivity, prognosis, and probable response to treatment [62, 63]. Biomarkers, whether for prognosis or prediction, are becoming more significant in personalized medicine to determine the subset of patients who will most likely to respond positively to a treatment [64]. However, the rigorous demand for reliable, consistent, and reproducible biomarkers has limited the application of precision medicine [65]. Prognostic biomarkers make it possible to classify individuals based on their likelihood of surviving their disease or advancing it. Thus, these can be used to modify the course of treatment for particular patients [66]. Although non-invasive biomarkers would have been preferable for classifying the patient and the disease, multiple invasive procedures—which are labor- and cost-intensive as well as draining for the patient are yet necessary in traditional practice to understand the development of a disease [67].

CAD systems integrated with AI employ several ML classifiers to assess clinical biomarkers such as lab results, patient history, proteomics data, and EEG signals and predict outcome [40, 68, 69]. Furthermore, with the implementation of DL in CAD systems the potential for discovering novel biomarkers has been demonstrated [18]. Anatomical information from imaging biomarkers plays a vital role in disease assessment. Radiological imaging biomarkers can distinguish between different stages of disease both spatially and over time, offering an advantage over fluid biomarkers due to their rich detail [70]. AI-CAD through the use of

deep learning (DL), now harnesses artificial neural networks to recognize patterns in large datasets and uncover hidden insights from complex data. This improves diagnostic accuracy and offers more reliable prognostic information, paving the way for personalized medicine and enhancing AI-CAD's ability to predict clinical events and disease outcomes [71]. The effectiveness of using AI-CAD for predictive and prognostic purposes has been established in a number of medical specialties, including oncology [72], renal damage [23], cardiology [28], diabetes [73], and neurology [74, 75].

2.3 AI Architectures in Prediction and Prognosis

2.3.1 ML Models

When it comes to early prediction or prognostic models in healthcare, ML classifiers have been crucial in their development. Several ML algorithms have proven to be particularly effective in creating models that could forecast treatment outcomes, patient risks, and disease outcomes by analyzing multiple features i.e., structured clinical data, laboratory findings, and basic imaging characteristics.

- Support Vector Machine (SVM) creates decision boundaries to classify data enabling the prediction of labels from one or more feature vectors [76] and has been used to predict survival in lung cancer patients [77] and neuroimaging risk factors in schizophrenia [78].

- Logistic Regression predicts a binary outcome, adjusting for multiple predictors and calculating how much each factor influences the result and adjusts for the effect of other factors [79]. It has been applied in predicting COVID-19 [80] and Alzheimer's outcomes[81].

- Decision Trees are popular for their interpretability and have been used in diabetes prognosis [82] and traumatic brain injury outcomes [83].

- Random Forest (RF), an ensemble method, combines multiple decision trees for accurate predictions, showed effective result in breast cancer prognosis model [84] and prognosis for elderly sepsis patients [85] .

- Multi-Layer Perceptron (MLP), inspired by the human nervous system, is well-suited for non-linear, real-world problems [86]. It has been used in prognosis models for prostate cancer [87], predicting intracerebral hemorrhage after stroke treatment [88]and for prognosticating renal cancer patients using histopathological images [89].

- k-Nearest Neighbors (KNN) classifies based on data similarity and has been applied in oral cancer staging [90] and cardiovascular disease outcomes predicting outcomes of ischemic cardiovascular diseases [91].

- XGBoost improves accuracy by combining predictions from multiple models and handles missing values well. It has been used to predict glioma patient outcomes [92] and prognosis for aneurysmal subarachnoid hemorrhage using clinical and demographic [93].

2.3.2 DL Models

Traditional ML classifiers have been proven to be effective for structured data, such as lab results and clinical records. However, to achieve image-driven medical predictions, DL has become essential for more complex predictive tasks, particularly when managing large amounts of unstructured data, such as radiological images. The term "deep" neural network refers to a neural network with multiple hidden layers. These layers can automatically learn using a general-purpose learning procedure; thus, the network is known as a deep learning network [58]. The functioning of the human brain inspires the architecture of these networks. By stacking several nonlinear processing layers, the original data is abstracted layer by layer. In each layer, various levels of abstract features are extracted from the data and utilized for object recognition, classification, or segmentation [94].

- **Convolution Neural Network**: Among the architectures proposed during deep learning's embryonic period, the one deep, feed-forward network that gained the highest popularity in the computer vision community was the convolutional neural network (CNN) [95]. It is perceived as a classic model that is excellent at extracting hierarchical features from images, making them ideal for diagnosing or predicting tasks involving large volumes of complex radiological scans.

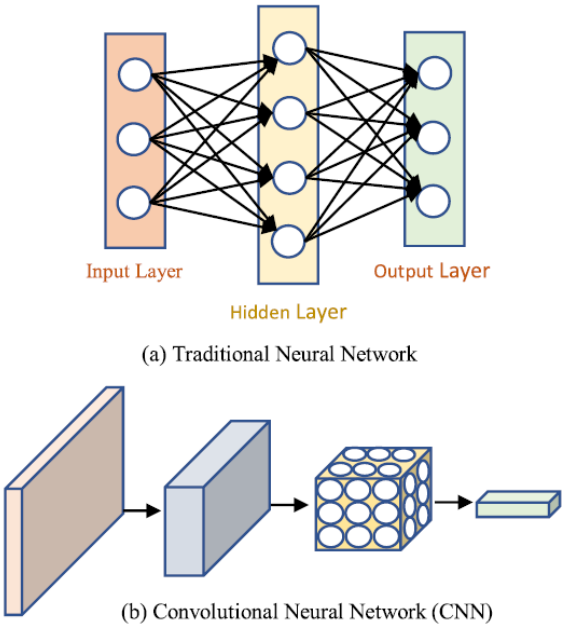

Figure 2.1: Traditional neural network vs convolutional neural network [2]

CNN is a deep model with supervised learning. It is a multi-layer perceptron that is quite similar to a traditional neural network. However, unlike a traditional NN, CNN takes an image as input (figure 2.1) and only connects in patches of the preceding layer rather than the entire layer [2]. As seen in figure 2.2 the structure of CNN consists of an input layer that has the same number of neurons as the pixels of the input image, a stack of hidden layers where each performs a definite operation to extract features (e.g. convolutional layer, non-linear activation layer, pooling layer and/or fully connected layer) and an output layer. An arbitrary size of the filter, also known as the kernel, works as a feature detector and produces a feature map or activation map by

convolving with the input image or the previous layer. The backpropagation technique is used to constantly optimize the kernel weights [3]. The established convolutional

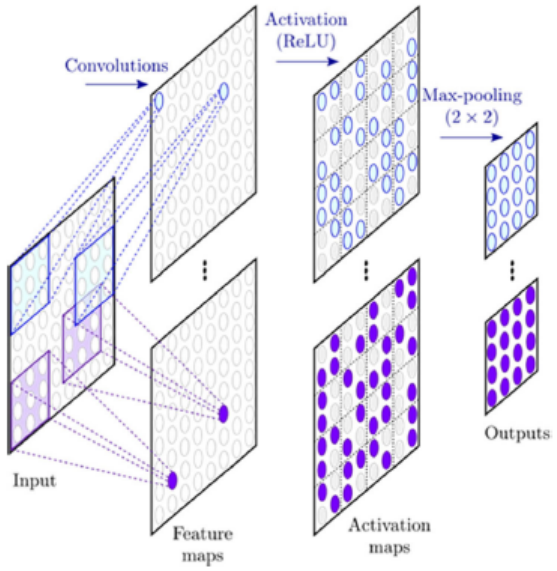

Figure 2.2: Structure of a CNN [3]

neural network is 2D where the input image and the convolutional kernel are both two-dimensional and have demonstrated predictive efficiency in fields like heart disease prediction [96] and Alzheimer's prediction using MRI [97]. However, to leverage the full potential of 3D scans, many 3D CNNs have been reported lately that can accept a volumetric image (from CT or MRI) as input and adopted as the prediction model for several diseases[98, 99].

- **Advanced CNN Architectures**: CNN architecture served as a skeleton for many advanced architectures that were built to overcome certain limitations or improve performance i.e., vanishing gradients, deep network efficiency, and feature reuse.

  – VGG is one of the earliest and most widely-used CNN architectures, gaining popularity since its success in the 2014 ImageNet competition. Its simple, consistent design, with layered convolutional filters, makes it effective for feature extraction, especially with limited datasets [100]. For example, Heidari et al. used the

VGG-16 model to predict COVID-19 outcomes based on chest CT images [101], while it has also shown strong performance in identifying Alzheimer's signatures and predicting the progression from cognitive impairment to Alzheimer's disease [102, 103].

- ResNet was developed to ease the training of a deeper network by introducing skip connection or residual connection, allowing the network to skip some layers to address the vanishing gradient issue[104]. In neuroimaging, ResNet has been used to predict Alzheimer's disease progression, outperforming traditional SVM models [105]. Additionally, it has been applied to forecast the prognosis of Diabetic Retinopathy, achieving an accuracy of 78.88% [106].
- DenseNet architecture also helps to avoid vanishing gradient but instead of residual connection, it uses dense connection to connect every other layer in the network in the network offering feature circulation while substantially reducing parameters to build efficient architectures [107]. DenseNet has been successfully applied in various medical prediction tasks, including predicting heart disease in cardiology [108], breast cancer from lymph nodes [109], and stroke patients' motor function using EEG signals [110]. These applications highlight DenseNet's effectiveness in handling complex medical data.

2.3.3 Hybrid Ensemble Model

To enhance prediction accuracy in AI-CAD, several studies have reported to development of hybrid models by combining ML and DL models which have yielded positive results. Radiomics features are straightforward to interpret, while deep features are more abstract, automatically learned by CNNs, and unique to each dataset. This provides versatility but also risks overfitting [111]. Hybrid models also incorporate ensemble techniques, where multiple classifiers are combined to improve performance. By leveraging the diversity of both feature types and ensemble methods, e.g., shown in figure 2.3, hybrid ensemble models can offer more robust and reliable predictions in AI-CAD systems. Ensemble techniques in

AI-CAD models involve combining multiple classifiers to improve predictive performance by aggregating their predictions.

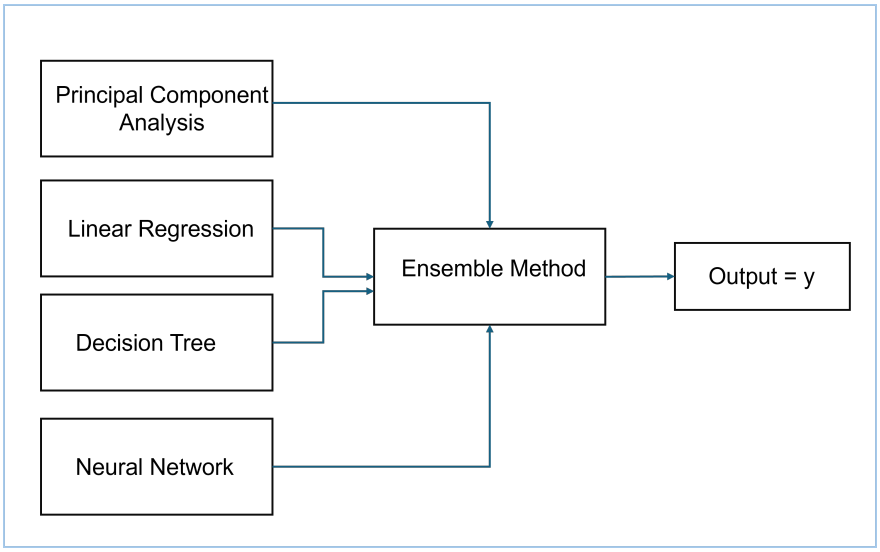

Figure 2.3: Hybrid ensemble model

The augmentation technique creates multiple variations of an image through data augmentation, which helps prevent overfitting and improves prediction accuracy by making the model less dependent on specific patterns. Stacking combines several machine learning or deep learning models to improve overall performance, making it useful for tackling complex tasks in CAD. Bagging trains multiple models with the same setup on different subsets of data, enhancing both accuracy and robustness. And lastly boosting builds a strong model by combining weaker ones, training each new model to fix the errors of the previous, creating a more accurate overall model [112]. Srinivas et al employed a hybrid model by stacking two CNN models and predicted COVID-19 patients using chest-CT [113]. Similarly, Venkatesh et al combined CNN with the Random Forest Algorithm to predict COVID using lung ct images [114]. Han et al. combined the features from deep transfer learning and radiomics and showed high performance in predicting the prognosis of patients with high-grade-gliomas [115]. Also, in the neuroimaging field, Lu et al combined deep features and radiomic features to predict hematoma expansion in intracerebral hemorrhage patients [116]. All these applications in different medical fields highlight the effectiveness of the hybrid model for

predicting disease or outcome.

2.4 Key techniques for optimizing AI models

As medical imaging data are complex with rich information, achieving a high-performance prediction model requires effective optimization techniques to learn proficiently from data and improve prediction accuracy.

- **Hyperparameters:** Hyperparameters are settings that control the learning process, influencing how the algorithm determines the optimal weights for the model's parameters. Some common examples of hyperparameters are as follows:

  - *Optimizer*: Optimizers are used to minimize loss function by updating the weights via a backpropagation algorithm. Stochastic Gradient Descent (SGD), Mini-Batch Gradient Descent, and Adam are among the popular optimizers. Choosing the best optimizer for the classification/prediction task based on the feature type or data size has high importance in model performance. Adam optimizer can adapt the learning rate with each epoch which makes it as preferred optimizer for predictive tasks [117, 118].

  - *Loss Function*: Loss function estimates the amount of error in weight updating and indicates whether the learning of the model is in the right direction. According to the target of the model task, one must choose among multiple standard loss functions [119]. Binary cross-entropy is the most used loss function for classification tasks and compares each of the predicted probabilities to the actual class output which can be either 0 or 1.

$$L_{BCE}(y, \hat{y}) = -(ylog(\hat{y} + (1-y)log((1-\hat{(}y)) \tag{2.1}$$

  - *Learning Rate*: The amount the weights are updated with respect to the loss function during training is referred as the learning rate; The rate often ranges between 0.0 and 1.0. It controls how quickly or slowly a model learns a feature.

- *Activation Function*: Activation functions are used to introduce nonlinearity to models which makes the model more powerful and adds the ability to learn complex features. It is responsible for transforming the summed weighted input from the node into the activation of the node or output for that input. Rectified Linear Unit(ReLU) is the most widely used activation function in the hidden layers [120]. Also, in the final layer Sigmoid or Softmax functions are preferred based on the number of class/es for the prediction task.
- *Dropout*: Dropout is a technique where randomly selected neurons are ignored or dropped during the training process. To optimize in what layer and how much weight should be dropped can assist the model in preventing over-fitting [121].
- *Number of Hidden Layers*: Increasing the hidden layers in between of input and output layer integrating with regularization techniques can increase the accuracy of the model [122].
- *Epoch*: The number of epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training dataset. The number of epochs is as high as possible until the loss saturates in the optimization process. Beyond this point, more epochs may result in overfitting the model.
- *Batch Size*: The batch size is a number of samples processed in training before the model is updated with new weights. It is critical as the computation of the loss function heavily depends on this hyperparameter.
- *Batch Normalization*: Normalizing the output of the previous (mostly activation) layer in mini-batches, to standardize before feeding to the next layer, by rescaling data to have a mean of zero and a standard deviation of one is known as batch normalization. It dramatically reduces the number of epochs which speeds up the training process [121].

All these hyperparameters can be optimized using random search or grid search opti-

mization techniques.

- **Data Augmentation:** In supervised learning, the evaluation of training heavily relies on the number of labeled data. Thus to increase the size of the training dataset data augmentation technique is applied. Data augmentation methods mainly include parametric transformation (e.g.: rotation, translation, shear, shift, flip). Some researchers also apply image enhancement tools (e.g.: noise suppression, changing image contrast and brightness) to add more data in the domain and add more insights into the model [123].

- **Transfer Learning:** Transfer learning refers to a system's ability to use knowledge gained from one domain to improve performance in a new task. Studies have shown that transfer learning often enhances model classification accuracy compared to random initialization. This process can be achieved at different levels, depending on how much of the pre-trained model is adapted, as shown in fig 2.4: full adaptation involves updating all network parameters during training, partial adaptation updates only the last few layers while freezing earlier ones, and zero adaptation uses the pre-trained model without any updates [124]. Choosing an appropriate transfer learning approach based on the similarity of the source and target tasks as well as the size of the target dataset can significantly boost the model's speed and accuracy. Improvised performance through transfer learning adaptation has been reported by several classification and prediction models (i.e., the Alzheimer classification model [125] and stroke prediction model[126].

- **Cross Validation:** The model's generalizability is assessed via a cross-validation technique. The cross-validation (CV) technique is applied to control the overfitting of the model and allows one to assess how well the model would perform on real-world data. Since limited training data is a prevailing issue, limiting the number of training samples and preserving them for testing becomes difficult. Thus CV technique groups the entire dataset into 3 parts: train, validation, and test; where test data is only to be

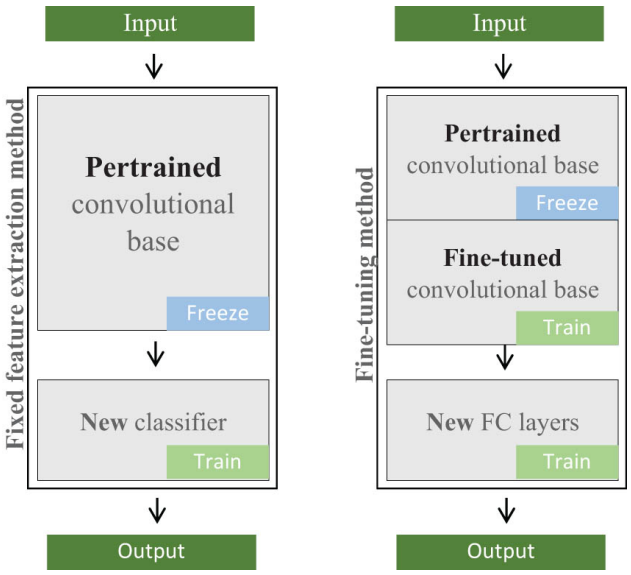

Figure 2.4: Transfer learning approaches: feature extraction and fine tuned [4]

utilized after the model is finalized, train data is for training the model, and validation data validates the performance of the model and indicates the requirement for any necessary changes. K-fold cross-validation is the most famous form of cross-validation technique, which divides the entire data, excluding test data, in K number of folds. K-1 folds are then fed into the training data for the model, which is subsequently verified on the fold that was left [127]. The process is repeated until each fold has been validated at least once. Other than this, Leave One OUT CV, which is more rigorous and leaves only one same in the validation data, and Stratified K-Fold CV, especially for imbalance class problems, are two other CV techniques applied [128]. CV techniques have shown their impact on model performance, such as in the early prediction of atherosclerosis and selecting features for cardiovascular disease prediction [129, 130]

- **Network Function Block** Different blocks or modules have been introduced in multiple studies for better learning performance of deep neural networks. A Dense Block connects all layers with matching feature-map sizes directly to each other and enhances the information flow. An attention block can selectively change input or assign differ-

ent weights to input variables based on their importance; e.g: a spatial attention block aims to calculate the feature importance of each pixel in the space domain and extract the key information of an image, and a channel attention block can feature recalibrate by using learned global information to accentuate selectively valuable aspects and reduce worthless features. These blocks are integrated into the network architecture and showcase better model performance in classification and prediction models (i.e., Prognosis after brain tumor, thorax disease classification) [131, 132].

- **Training Framework** Aside from high-performance GPUs, which are attributed to the prosperity of deep learning networks, the development of a number of software frameworks based on deep learning theory is also bringing convenience and speed to a plethora of DL tasks. *Pytorch*, *Theano*, *Tensorflow* are some popular frameworks for deep learning. Among these, *Tensorflow*, which is an open-source software library by Google, is the most popular among researchers. The popularity is mostly due to its high-level API, *Keras*, which is simple and beginner-friendly. Google also offers a free online cloud-based environment called *Google Colaboratory* that allows the training of deep learning models on CPUs, GPUs, and TPUs.

- **Network Performance Metrics** The performance efficacy of the prediction model is assessed using standardized and well-known metrics, allowing the system to be compared to other approaches in the literature. The model target operation determines the selection of a suitable evaluation metric. Computing complexity, processing time, memory consumption, and accuracy are some metrics that may be measured with these measures. Medical image classification tasks are primarily evaluated using metrics that rely on the spatial overlap between the two compared classes. This evaluation is based on four fundamental overlap cardinalities: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The total number of accurately classified instances is calculated by adding TP and TN, while the total number of incorrectly classified instances is determined by summing FP and FN.

- *Accuracy*: The percentage of accurately classifying image pixels is referred to as overall pixel accuracy, shortly known as accuracy metrics. It is the most fundamental performance metric [133].
- *Area Under Curve*: The area under the receiver operating characteristic curve (AUC) is a commonly used metric for assessing the discriminatory power of a clinical prediction mode interpreted for specific thresholds. For each threshold, the True Positive Rate (TPR) and False Positive Rate (FPR) are calculated and plotted on a graph. Ideally, a higher TPR and a lower FPR indicate better model performance [133].
- *Sensitivity*: It quantifies a model's ability to identify patients with positive outcomes by measuring the proportion of subjects with an actual positive outcome who are correctly classified as positive [134]. It is also known as 'Recall'.
- *Specificity*: It measures how well a model identifies patients without the condition by showing the percentage of those who are correctly classified as negative.
- *Precision / Positive Predictive Value*: In a disease prediction model, precision refers to its ability to accurately identify only the relevant data points minimizing the false positive rate[133].
- *F1-Score*: F1 is the harmonic mean of precision and recall, which weighs both equally. It combines these two metrics to evaluate how well a model accurately identifies positive outcomes in patients while reducing false positives and false negatives [135].

Optimizing AI models has been essential across various medical fields. Stroke, being a major health issue with significant social and economic impacts, has become a key focus for researchers. They use advanced AI-CAD systems to improve stroke prognosis and prediction, aiming to enhance accuracy and treatment efficiency. These optimization strategies are crucial for developing reliable prognostic models and improving the accuracy of stroke outcome predictions by combining clinical and imaging data effectively.

2.5 Stroke Prognosis and Prediction Model

In 2022, the World Stroke Organization identified stroke as the second leading cause of death worldwide and the third leading contributor to combined death and disability, measured as disability-adjusted life years (DALYs). From 1990 to 2019, the global burden of stroke significantly increased, with a 70% rise in new stroke cases, 43% more deaths, 102% more prevalent cases, and a 143% rise in DALYs [136]. Currently, about 34% of global healthcare spending goes towards stroke. In the United States, the average cost per stroke patient—including hospital care, rehabilitation, and follow-up—totals around USD 140,048 [137]. Thus the need for accurate and timely prediction tools became crucial due to the significant impact that chronic illnesses have on patients, families, and healthcare systems [138], As, AI-CAD has emerged in optimization technologies, that advancement has well reflected showing high potential in the stroke prediction [139].

Strokes are primarily divided into two types: ischemic stroke and hemorrhagic stroke. Hemorrhagic stroke occurs when a blood vessel in the brain ruptures, leading to bleeding into brain tissue. This type of stroke, often caused by an intracranial aneurysm, accounts for about 13% of all stroke cases. Ischemic strokes, which make up 87% of stroke cases, occur due to blockage of blood to brain regions and suddenly impairing function [140].

2.5.1 Hemorrhagic Stroke

2.5.1.1 Background and Prognostic Factors

aSAH is a type of stroke with high mortality and morbidity rates, affecting nearly 30,000 individuals annually in the United States. This condition arises from the rupture of an intracranial aneurysm, leading to bleeding in the subarachnoid space, specifically in the cisternal spaces of brains, that is, the area between the brain's surface and the thin tissues covering it [46]. The severity of this medical emergency is underscored by a considerable proportion of patients experiencing poor clinical outcomes, with an in-hospital case-fatality rate of approximately 30% [141]. Given the high percentage of patients experiencing poor

clinical outcomes, predicting the short-term prognosis is vital for effective in-hospital treatment planning and avoiding unnecessary interventions. Additionally, standardized follow-up periods ranging from 3 months to 1 year can pose risks for patients requiring closer attention and can be cost-prohibitive for those with better prognoses. Accurately anticipating long-term clinical outcomes can address both issues by optimizing patient care and resource allocation.

Studies indicate that critical indicators for the prognosis of aneurysmal aSAH include age, clinical, and radiological characteristics at admission e.g., clinical grade(i.e., Hunt and Hess [H&H] score, World Federation of Neurosurgical Societies [WFNS] score), radiological grade (i.e., Fisher or modified Fisher score [mFS]), aneurysm dimensions, hemorrhage volume, and reduced sulcal volume [142]. Additionally, Early Brain Injury (EBI)(i.e., events during the first 3 days of aSAH) and Delayed Cerebral Ischemia (DCI)(i.e., events during 4-14 days of aSAH) are significant predictors of mortality and morbidity in aSAH patients. EBI refers to the immediate cellular, chemical, and physiological changes that occur within approximately 72 hours after the hemorrhagic event, resulting from the post-bleeding effects in the subarachnoid spaces. Factors contributing to EBI include the extravasation of hemorrhagic blood, a sharp increase in intracranial pressure, the sudden onset of arterial vasospasm, decreased cerebral blood flow, disrupted cerebral autoregulation, and brain swelling. These factors collectively contribute to the severity of EBI and impact the overall prognosis of aSAH patients [143].

EBI can set the stage for the development of DCI, a potential clinical complication that typically manifests within a few days to a week after the initial hemorrhage and is the leading cause of death and disability following aSAH. It is characterized by a reduction in blood flow to specific brain regions [144]. EBI can also lead to Clinical Vasospasm (CVSM), the constriction of blood vessels, resulting in reduced blood flow, ischemia, and further neurological impairments [145]. Another common outcome of aSAH is hydrocephalus (HCP), which is caused by a blockage in the normal cerebrospinal fluid pathways due to blood breakdown [146]. Long-term impacts of aSAH can include functional impairment, mood

fluctuations, personality changes, and emotional lability. Two well-known indicators are usually utilized to monitor the patient's long-term prognosis: the modified Rankin Scale (mRS) to measure physical disability and functional status, and the Montreal Cognitive Assessment (MoCA) to assess cognitive abilities.

2.5.1.2 Prognostic Models and Gaps

Diagnosing DCI after aSAH onset can be difficult, especially in sedated patients, as there's no reliable test for it. Traditional approaches to treating aSAH usually focus on managing narrowed arteries to control EBI and DCI [141]. Vasospasm, the event of sudden artery narrowing, often gets linked to DCI, but not all patients with vasospasm get DCI and vice versa [147]. Early detection using cerebral perfusion CT scans (CTP) showed promise, but standardized thresholds for predicting DCI remained inconsistent due to equipment and method differences [148]. Despite the development of various grading scales for predicting the clinical events and outcome of SAH, their accuracy, and generalizability are limited, reducing their clinicalusefulness[149, 150]. Risk factors, clinical, and radiological scales have historically been used to predict mortality and outcomes, but these models haven't been very effective in practice and lack strong external validation [151].

Many studies have applied AI-CAD systems in developing the outcome prediction model. Katsuki et al. employed 24 characteristics including demographic information (i.e., age, sex, hypertension, etc.), clinical variables (i.e, hemoglobin A1C, albumin, cholesterol, etc.), and radiological findings (i.e., aneurysm location, size, temporal muscle thickness, etc.) to develop Auto-AI based prediction model for DCI and functional outcome. While the AI model outperformed these traditional methods for functional outcomes (AUC = 0.801), it did not significantly outperform for predicting DCI (AUC = 0.65) [152]. Lo et al. used a large database of demographic and clinical variables in a CNN with fuzzy logic to achieve an AUC of 0.85. But as several features during admission were missing, the model was not feasible to use for early prediction [153]. Dengler et al. compared machine learning (ML) methods for predicting outcomes with clinical and radiographic data in aSAH patients and

found that traditional scores performed similarly to ML models, with Glasgow Coma Score (GCS) and age being key predictors [154]. Toledo et al. developed a machine learning-based decision tree model using the Fisher and WFNS scales, achieving an area under the ROC curve (AUC) of 0.85 [155].

Several other clinical biomarkers (e.g., Blood components, Cytokines, MicroRNAs, C-reactive protein (CRP), Glucose, and S100B) have also been explored to study the effects of aSAH prognosis [47]. All these studies employed demographic data, clinical biomarkers, and radiological features as factors to predict the outcome of aSAH, but not many research efforts were reported to examine the efficacy of radiomic image-based biomarkers for developing the prognostic model. Isabel et al. discussed the significance of neuroimaging data for the prediction of the outcome and how the initial finding in the radiological scans can assist in prognostic understanding [156]. Garcia et al. employed initial baseline CT with the CNN model to predict the mortality of patients with AUC score 0f 0.80 [157]. On the other hand, several studies have highlighted subarachnoid cisternal blood volume as one of the main crucial factors [158, 159]. The detection and measurement of blood volume pose significant challenges due to potential inter-rater and intra-rater variability. A semi-automated qualitative approach using a modified Fisher scale (mFS) to assess subarachnoid and ventricular blood has proved promising in predicting DCI. Still, it remains subjective and exhibits high inter-rater variability [145]. Similarly, the quantification of blood using acute HCP, transient cerebral oligemia, and systemic inflammatory response has a reasonable capacity to predict the onset of DCI and long-term clinical outcomes; however, it is also semi-automated and dependent on the operator[160].

Despite the significant role of hemorrhagic blood in prognosis, there has been limited research on quantifying blood volume as an imaging biomarker, hindering its use in evaluating disease severity and prognosis. Further research could assess the efficacy of deep learning and/or hybrid models, discussed for a range of diseases in sec 2.3, in developing prognostic tools for aSAH patients.

2.5.2 Ischemic Stroke

2.5.2.1 Background and Prognostic Factor

While aSAH is caused by blood leaking into the brain and leading to bleeding, AIS occurs due to blood flow blockage in a brain region, causing tissue damage. Over 7.6 million new AIS cases are reported annually [136]. AIS patients' functional outcome depends more on their initial clinical presentation and radiological findings than on treatment strategies [161]. Early ischemic changes visible on radiology scans can indicate potentially irreversible damage. Earlier, the European Cooperative Acute Stroke Study (ECASS) trials revealed that prognosis depends on the extent of middle cerebral artery (MCA) involvement, but accurate identification remains difficult even for experienced professionals [162].

To address this limitation, the Alberta Stroke Program Early CT Score (ASPECTS) was developed as a robust imaging measure for predicting functional outcomes and symptomatic hemorrhage following iv-thrombolysis and has since been widely used in clinical practice to evaluate early ischemic changes in the MCA territory [52]. ASPECTS offers a systematic and reliable way to assess early ischemic changes on a CT scan. Originally developed for non-contrast CT scans, this 10-point topographic scoring system has been extended to include DWI-MRI scans, offering greater sensitivity. It divides the MCA territory into 10 regions as shown in fig. 2.5, assigning 1 points for each region with normal MCA appearance and deducting points for affected regions. Subsequently, recognizing the prevalence and unique challenges of posterior circulation infarctions, the scoring system was extended to the posterior area, and hence PC-ASPECTS was introduced as a parallel scoring system [163]. The 20-point scoring (10 points ASPECTS and 10 points PC-ASPECTS) has proved to better consistency and agreement between observers compared to the older "one-third MCA rule," which required clinicians to estimate whether more than one-third of the middle cerebral artery (MCA) territory was affected by ischemia.

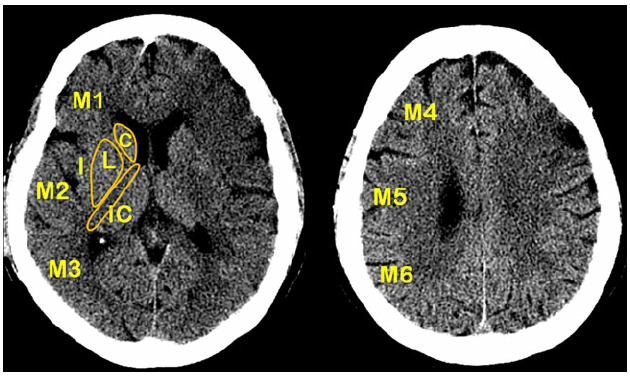

Figure 2.5: ASPECTS area in CT imaging[5]

2.5.2.2 Prognostic Model and Gaps

Early studies suggested that patients with ASPECTS score< 7 were identified as higher-risk patients and critical candidates for mechanical thrombectomy over the patients with lower scores [75]. Despite the widespread popularity of ASPECTS in clinical practice due to its simplicity, conflicting results also have emerged regarding the inter- and intra-observer variability of ASPECTS[164, 165]. Good agreement is seen for patients with higher ASPECTS scores(i.e.,> 7), but significantly lower performance is observed for patients with low ASPECTS scores with little ischemic changes. Additionally, research reveals that not all ASPECTS regions hold the same prognostic value for predicting the functional score [166], yet the traditional ASPECTS approach weights all regions equally, which limits predictive precision [167]. Recent advancements in AI and ML are transforming stroke diagnosis and prognosis. Studies in AI and ML applications for medical imaging highlight that these technologies can significantly improve the accuracy and reliability of stroke assessments by automating the analysis of complex imaging data, making diagnoses faster and more consistent [168].

SVM and neural networks were used to predict the 90-day mRS outcome in a study involving 107 patients with acute anterior circulation ischemic stroke who received endovascular therapy. The neural network achieved a test AUC of 0.6. The authors reported that SVM showed mixed performance in predicting specific mRS scores, which accurately pre-

dicted outcomes with 100% accuracy for patients with a 90-day mRS score of 6 but only 25% accuracy for those with a score of 3 [169]. Another machine learning model was developed by Lin et al. using clinical, functional, and quality of life assessment scores as input features to predict the Barthel Index (BI)—an indicator of basic independence for post-stroke patients. The model showed promising accuracy in predicting patients' ability to perform activities of daily living (ADL) at discharge [170]. Zhou et al. conducted a logistic regression analysis comparing radiomics, clinical, and combined clinical-radiomics models. Their findings showed that the combined clinical-radiomics model performed best, achieving an AUC score of 0.87 for predicting six-month outcomes in stroke patients [171]. Oliveira et al reported a hybrid model developed by combining imaging biomarkers extracted by a 3D-CNN model with clinical data. The findings in this study underscore the potential for deep learning models that incorporate both imaging biomarkers and clinical data to improve the accuracy of outcome predictions for ischemic stroke patients, though further refinement was needed for clinical application [45]

Besides clinical data and radiological scores, recent studies have established ischemic volume as a key factor for outcome prediction[172]. But to account for variations in brain size especially since brain volume can differ significantly among individuals(i.e., infant) it is often more accurate to express ischemic volume as a percentage of total brain volume [173]. Subsequent studies have shown that ASPECTS blood volume is a more effective imaging biomarker for predicting stroke outcomes than the traditional ASPECTS system [174]. However, the ASPECTS volume regions are not uniform, and losing a small point can represent different amounts of ischemic damage depending on the affected area. Some regions are linked to a higher risk of poor outcomes, indicating that damage in those critical areas can more significantly impact long-term prognosis [55]. On the other hand, advances in AI and ML are proven to enhance stroke diagnosis and prognosis by automating the analysis of complex imaging data, improving precision and consistency. Thus integrating these technologies will address limitations in subjective assessments, leading to more accurate and reliable predictions of patient outcomes.

Evaluating ASPECTS (and PC-ASPECTS) volumes as a percentage of total ischemic volume may enhance predictive accuracy, yet few studies have quantitatively analyzed volumetric biomarkers for ischemic stroke prognosis from this perspective.

In summary, current prognostic models for both types of stroke have limitations that advanced AI technologies can help address. For SAH, an optimized prognostic model can be achieved by utilizing an automated quantitative imaging biomarker based on deep features or hybrid ensemble features. For AIS, accurate prognosis can be obtained through region-based volumetric analysis. Adapting and testing various AI-CAD models, along with applying the optimization techniques discussed in this chapter, has proven useful in developing these prognostic models. Detailed discussions of these developments will follow in the next three chapters.

CHAPTER 3

Evaluation of deep learning framework coupled with interactive user interface to predict clinical complications after Aneurysmal Subarachnoid Hemorrhage

3.1 Introduction

aSAH is a type of hemorrhagic stroke and a medical emergency with high rates of death and disability, affecting around 30,000 people each year in the United States. When an aneurysm ruptures, blood leaks into areas around the brain called cisterns. The fatality rate is about 30%, and many survivors are left with physical and cognitive impairments. [175]. Previous studies have identified several prognostic factors for aSAH, including age, clinical and radiological characteristics at admission, radiological score (mFS), hemorrhagic blood volume, and reduced sulcal volume [142].

EBI and DCI are key factors that lead to complications in aSAH survivors. EBI includes immediate cellular, chemical, and bodily changes that happen within the first 72 hours after the hemorrhage. These changes include hemorrhagic blood extravasation, increased

intracranial pressure, transient vasospasm, reduced cerebral blood flow, impaired cerebral autoregulation, and brain swelling [143]. The risk of DCI depends on the amount of blood from the initial hemorrhage; more blood and its breakdown products disrupt brain function, damage blood vessels, and cause small clots through inflammation[144, 176]. Additionally, hydrocephalus (HCP), characterized by increased cerebrospinal fluid (CSF) in the ventricles and subarachnoid spaces, contributes to EBI and may result in prolonged cognitive effects among survivors[146]. Long-term outcomes in aSAH patients include functional impairment, mood fluctuations, personality changes, and emotional liability. The mRS and MoCA are widely accepted clinical tools for evaluating short-term and long-term physical and cognitive outcomes.

Although EBI has been recognized as a crucial factor influencing poor functional and cognitive outcomes in aSAH patients, it has only recently begun to receive significant attention. Traditionally, most studies have focused on preventing and treating arterial narrowing, known as vasospasm to address DCI [141]. The timely and accurate prediction of patient outcomes following aSAH is essential for informed clinical decision-making and for providing comprehensive information to patients and their families. Furthermore, it is critical to plan personalized patient visits and determine customized follow-up durations rather than adhere to a standardized follow-up routine for all patients. However, it is difficult for clinicians to predict these short- and long-term outcomes due to variations in clinical presentation and the diverse courses of disease progression throughout hospitalization. Therefore, developing a method for early, objective, personalized outcome prediction in aSAH patients can help with result forecasting, resource allocation, and individualized follow-ups.

Numerous studies have tried to assess SAH outcomes by integrating clinical, physiological, pathological, and radiological variables; nevertheless, they lack independent external validation [149]. Several clinical biomarkers i.e. C-reactive protein, S100B, Glucose, Neuroglobin, and others have been studied as predictive for aSAH patient outcomes [47]. Despite the crucial significance of imaging biomarkers discussed in section2.2, limited studies have explored the potential of imaging biomarkers as predictive tools for aSAH prognosis. To

address post-aSAH complications, a CAD system was developed in the pilot project of this study, primarily focusing on generating radiographic image markers related to sulcal volume. The study's findings were promising, demonstrating the system's potential to predict clinical outcomes in aSAH patients both in the short and long term outperforming models based on other clinical biomarkers [177]. Multiple studies have reported that predicting outcomes can be accurately achieved by assessing the hemorrhagic blood volume in specific areas of the brain post-asAH[159, 46]. A semi-automated qualitative approach using a modified Fisher scale (mFS) has been proposed to assess subarachnoid and ventricular blood and has proved promising in predicting outcomes. Still, the approach remained subjective and exhibited high inter-rater variability [178].

Building on these efforts, my first study aims to develop an imaging biomarker based on hemorrhagic blood volume. This work proposes the development of an automated CAD system designed to quantify various hemorrhagic characteristics from CT images aquaried on different time period to forecast prognosis following an acute subarachnoid hemorrhage (aSAH). This objective was accomplished by employing advanced deep-learning architectures. Since 2012, Convolutional Neural Networks (CNNs) have shown significant performance improvements and are now extensively used in various image classification tasks. Notably, CNNs have even surpassed human experts in certain studies involving medical image classification [179]. Numerous studies have demonstrated the effectiveness of deep learning architectures in various modalities for the classification and prediction of diverse diseases [113, 132, 180, 181]. Therefore, the purpose of this study was to design and assess the effectiveness of deep learning frameworks to categorize patients' chances of developing a spectrum of short- and long-term clinical issues as a result of aSAH. No previous studies were reported using a deep learning-based classification architecture to treat these post-aSAH clinical situations.

3.2 Methodology

A dataset of non-contrast CT images paired with clinical data was assembled for this purpose. Various deep learning architectures were employed, and transfer learning was utilized to develop the initial prognostic model. Additionally, an attention mechanism module was integrated to enhance focus on relevant features, improving prediction accuracy. This section provides a comprehensive overview of these methods, along with the performance evaluation metrics used.

3.2.1 Dataset

The Institutional Review Board of the University of Oklahoma Health Sciences Center (OUHSC) approved this study, ensuring all participants had consented to engage in the extended follow-up phase, as confirmed during their one-year clinical review. This research tracked 60 patients who were diagnosed with aSAH at OUHSC from 2014 to 2016, necessitating both an aSAH diagnosis and a CT scan conducted within 48 hours from the onset of symptoms for study inclusion. Throughout their hospitalization, patients underwent multiple CT scans, starting with an admission scan followed by several interim scans to monitor the progression of their condition. However, for this study, attention was focused solely on analyzing the first admission CT scan and the discharge CT scan taken 10 to 14 days after admission, a period specifically chosen to coincide with the likely conclusion of the DCI phase, thus aligning with the timing of patient discharge. These pivotal scans are analyzed through a CAD scheme for classification purposes, with the outcomes subsequently reviewed by the doctor for a comprehensive evaluation.

Figure 3.1 illustrates the current clinical practices, their shortcomings, and the motive of our study in evaluating patients with aSAH. Specifically, it highlights the challenges of emerging concurrent issues during hospitalization, routine follow-ups, and the scope of developing predictive ensemble models to address such issues to develop proactive and personalized CAD systems.

In the retrospective study,8 clinical complications were closely monitored. Among them,

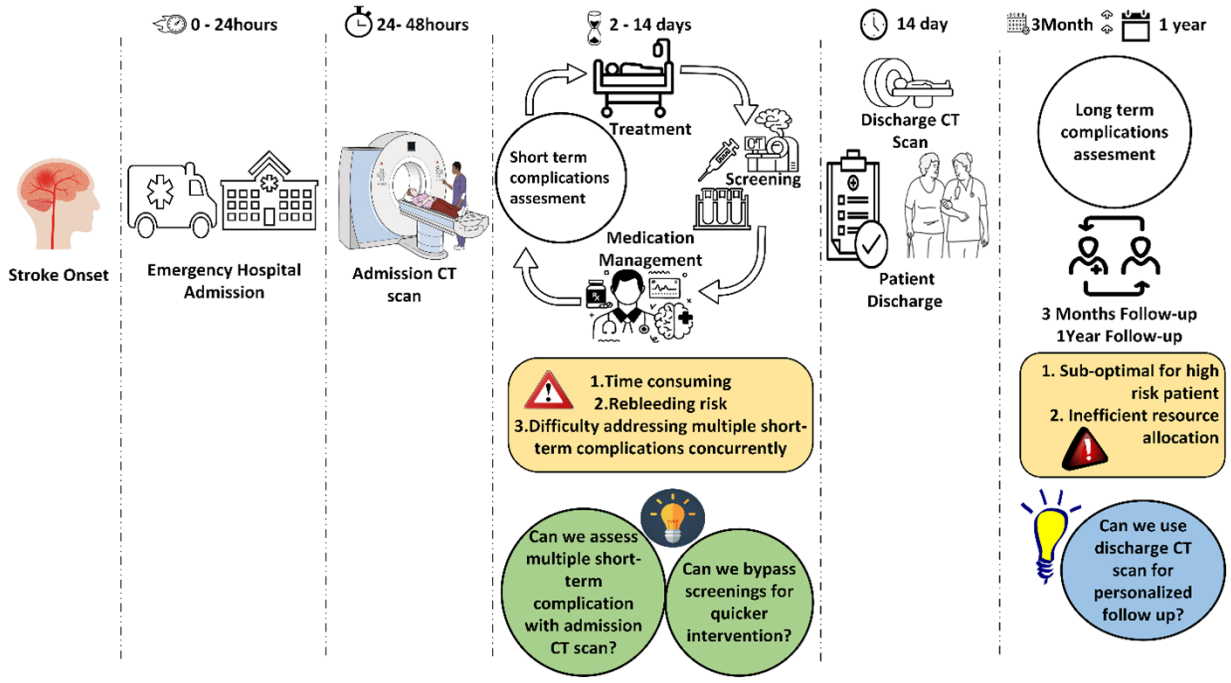

Figure 3.1: Patient evaluation and diagnostic process

4 short-term complications included clinical vasospasm (CVSM), clinical stroke-like symptoms associated with narrowing intracranial blood vessels; DCI, identified by brain edema in the scans from blocked blood flow; HCP, involves an accumulation of cerebrospinal fluid in the brain, identified through targeted measurements early in the patient's care and ventriculoperitoneal shunt (VPS), a surgical procedure to relieve brain pressure by redirecting excess fluid. VPS procedures can occur either during the patient's hospitalization or as a follow-up intervention after discharge. These short-term outcomes were categorized into two classes based on the absence/'class1' or presence/'class2' of each outcome.

Long-term outcomes are typically assessed through post-discharge evaluations highlighting the physical and cognitive functions of an aSAH patient. These outcomes were measured with two key indices: the mRS, which ranges from 0 (no disability) to 6 (severe disability), to assess physical disability, and the MoCA, scoring between 0 to 30, where lower scores indicate cognitive impairment. These measures provided a comprehensive view of the patient's functional status following discharge, highlighting the spectrum of potential long-term conse-

Table 3.1: Clinical data with demographic information of patients

	Admission Scan				Discharge Scan			
	Class 1		Class 2		Class 1		Class 2	
Clinical Outcome	Age	M/F (N)	Age	M/F (N)	Age	M/F (N)	Age	M/F (N)
DCI	52.92	12/13	52.23	9/15	52.96	12/12	54.14	6/22
CVSM	53.94	13/23	54.20	6/9	51.75	6/18	54.38	5/8
HCP	51.52	6/19	51.78	8/15	54.24	12/21	52.46	6/13
VPS	51.74	14/25	53.64	4/9	51.45	12/21	55.75	3/9
mRS 3M	52.6	9/16	52.07	5/10	52.86	8/13	59.15	3/10
MOCA 3M	55.52	6/17	57.47	5/10	56.65	5/12	58.46	3/10
mRS1FU	55.62	6/15	53.47	8/9	56.63	4/15	53.53	7/8
MOCA 1 FU	57.9	6/14	54.4	5/10	61.71	3/4	56.14	5/17

quences stemming from their initial clinical presentations. Outcomes were assessed through post-discharge evaluations at 3-month (3M) and 1-year (1Y) follow-ups. In this study, mRS scores of 0 to 2 were categorized as good outcomes (independent functioning)/'class 1', while scores of 3 to 6 indicated poor outcomes (dependent functioning or death)/'class2'. Similarly, MoCA scores of 26 or higher were classified as good (normal cognitive function)/'class1', and scores below 26 as poor (cognitive impairment)/'class2'. Table 3.1 provides a summary of the patient's demographic data and events studied.

3.2.2 Image Processing and Data Curation

The presence of substantial subarachnoid clots on the initial CT scan upon admission has been correlated with the subsequent occurrence of DCI post-aSAH. Existing evidence suggests that the mechanical or natural removal of cisternal blood clots is associated with a reduced risk of vasospasm [145]. Consequently, the primary factor guiding the objective of our study and the selection of image data is the blood volume in the cisternal area.

First, to segment raw brain CT data, a two-stage consecutive mapping-based algorithm

was deployed. Initially, preprocessing was conducted on each 3D CT scan sequence to detect three image markers: global minimum (gm), local minimum (lm), and maximum peak (mp) within the intracranial brain region. These markers served as crucial reference points guiding the segmentation algorithm. Following this, Otsu thresholding was applied to isolate the largest single-connected intracranial brain region in each 2D-image slice to segmented brain volumes. Further refinement was then executed to enhance accuracy and address leakage issues in the segmentation process. More details about the segmentation algorithm are available in our previous study [177]. The segmented brain area was then further CT labeled into four distinctive regions corresponding to the intensity threshold of the HU values, which were labeled as follows: those with HU values less than or equal to 20 were categorized as cerebrospinal fluid (CSF), values exceeding 32 were identified as blood, those falling between 20 and 32 were labeled as white matter (WM), and values between 32 and 60 were designated as gray matter (GM).

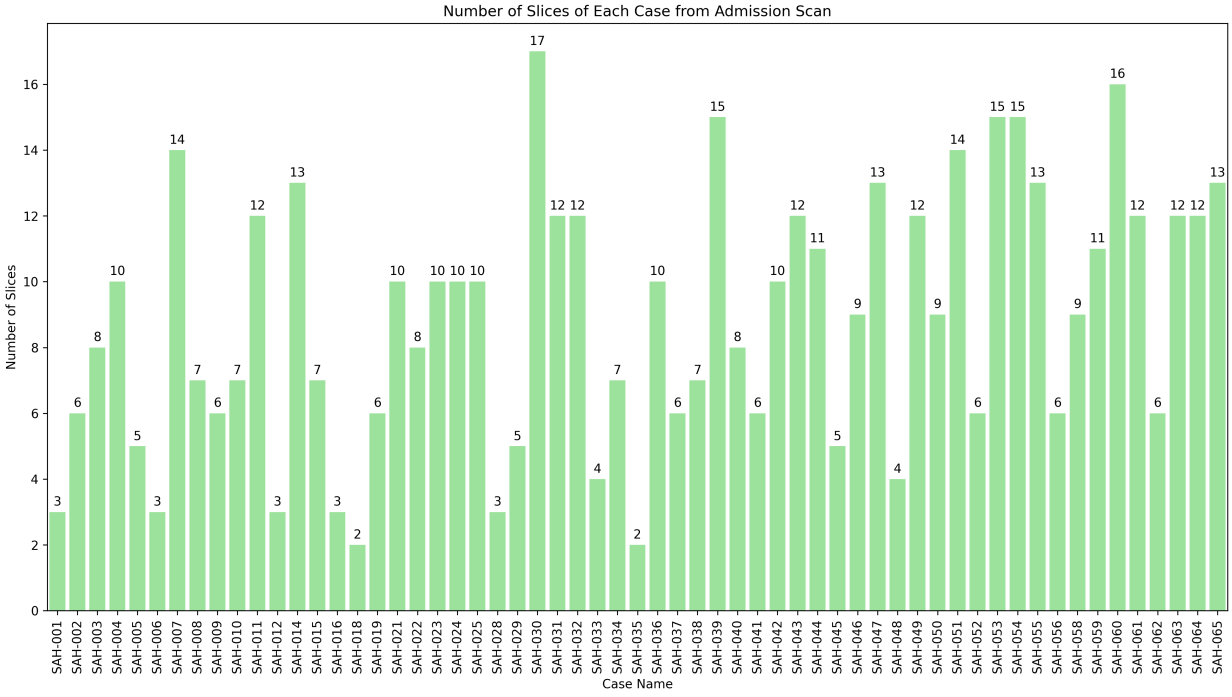

Figure 3.2: Number of slices of each case from admission scan

As mentioned earlier, the presence of significant subarachnoid blood clots on the initial

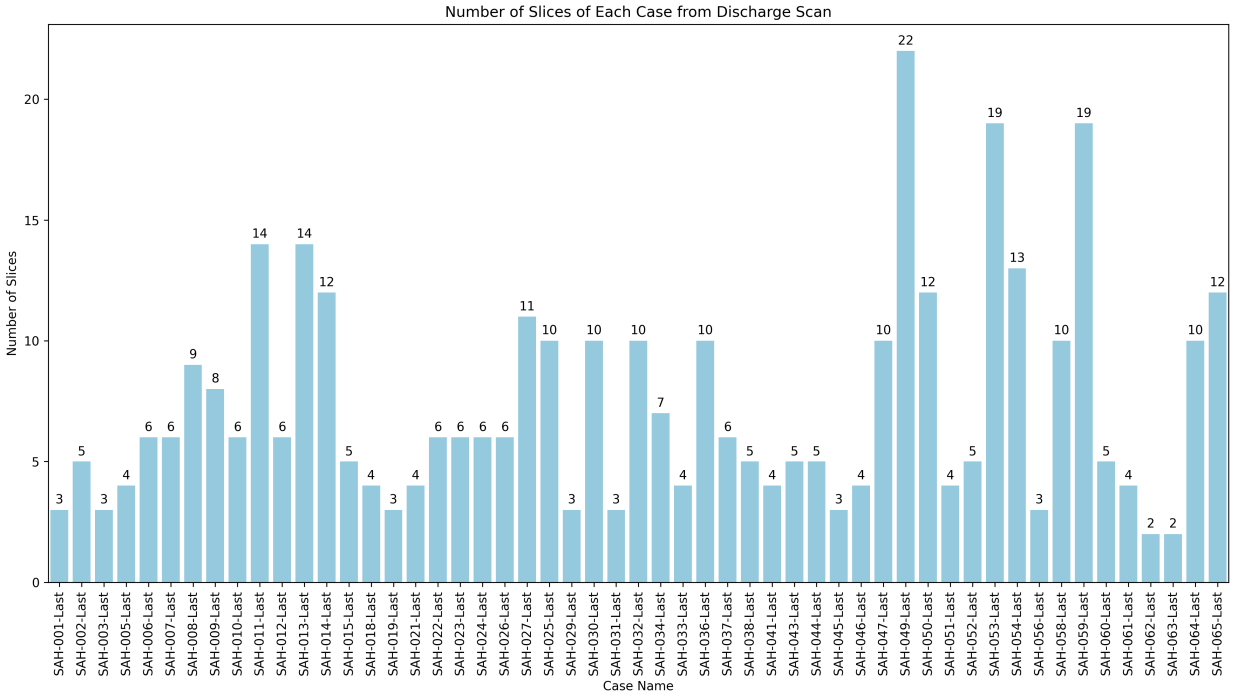

Figure 3.3: Number of slices of each case from discharge scan

CT scan at admission has shown a correlation with the subsequent development of DCI and outcomes following aSAH. Prior studies indicate that mechanical or spontaneous removal of the cisternal blood clots is linked to a decreased risk of EBI and future outcomes [145]. Hence, the principal determinant driving my study's objective and the choice of image data is the blood volume present in the cisternal region. After segmenting the brain into labeled regions, CT slices from each case were selected that featured noticeable blood clots in the cisternal space, marking the corresponding blood regions. Each patient's labeled CT slices containing cisternal blood information and their original and segmented image slices were exported to the local directory for further processing. The number of slices varied depending on the presence of blood in relevant areas. Figure 3.2 and 3.3 highlight the number of slices extracted from each case both for the admission scan and discharge scan respectively.

In the slice count analysis for each outcome, distinct values were observed between admission and discharge scans for Class 1 and Class 2. Generally, patients with a higher number of blood-marked slices tended to fall into Class 2. Figures 3.4 illustrate examples of short-term

(DCI) and long-term (mRS3M) outcomes, respectively, highlighting the slice counts for each case. The blue bars imply cases with class 1, while the orange bars signify the class 2 cases for both outcomes. Table 3.2 summarizes the slice counts for all outcomes by class across both scan periods.

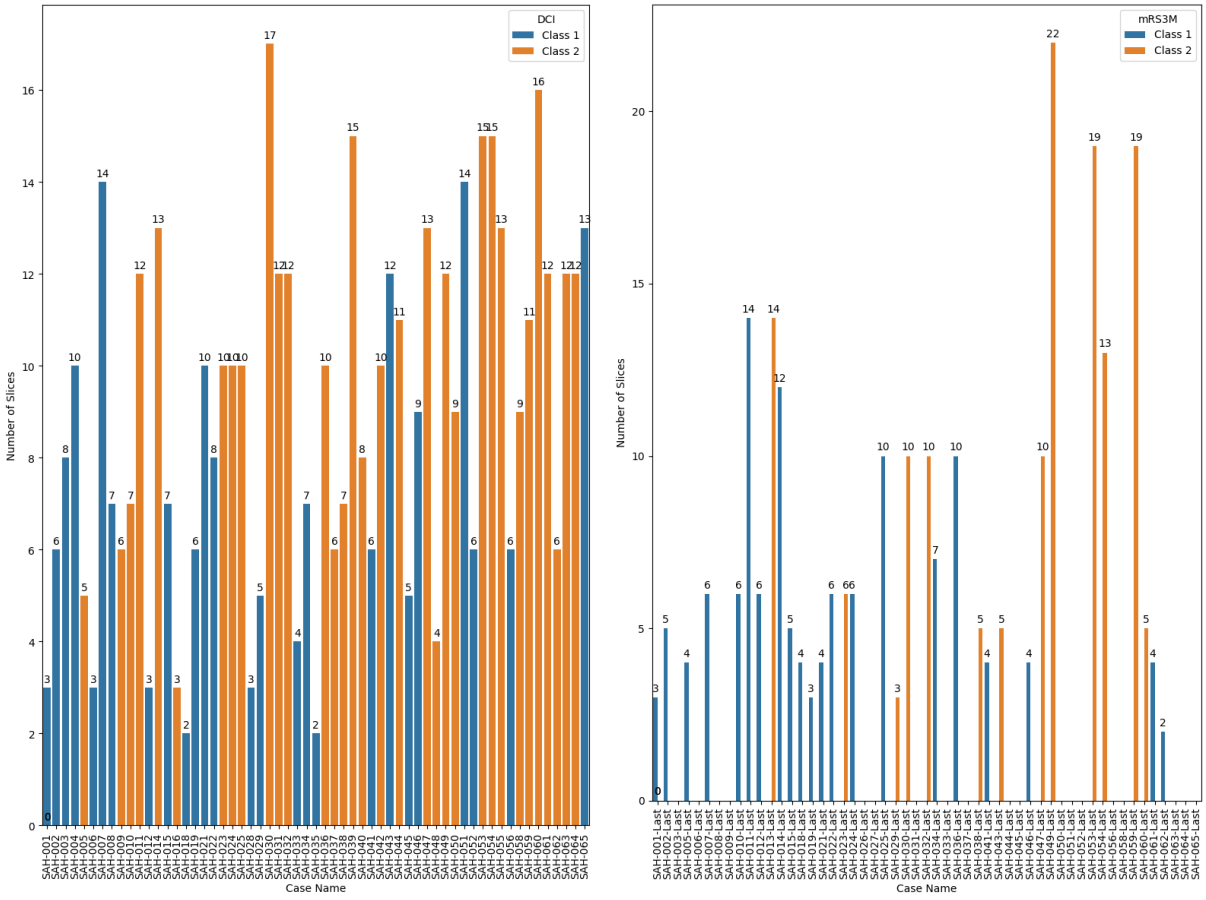

Figure 3.4: Number of slices for each case: (left) DCI (short-term) and (right) mRS3M (long-term)

A five-fold approach was adopted to optimize the quantity of training data and suitably assess the efficacy of the predictive models because of the small dataset size. To guarantee that there is no information leakage between the folds and also that the amount of samples in each fold is fairly balanced, appropriate design consideration was followed. Both a base

Table 3.2: Number of cases and slices by outcome and class across admission and discharge scan

Outcome	**Admission Scan**				**Discharge Scan**			
	# of Cases		# of Slices		# of Cases		# of Slices	
	Class 1	**Class 2**	**Class 1**	**Class 2**	**Class 1**	**Class 2**	**Class 1**	**Class 2**
DCI	26	33	179	343	24	28	120	254
CVSM	36	23	282	240	33	19	196	178
HCP	24	15	197	135	24	13	179	86
VPS	40	13	319	130	33	12	200	94
mRS3M	27	14	205	170	21	13	125	141
MOCA3M	19	22	160	215	16	18	102	164
mRS1F	22	14	198	151	17	13	119	121
MOCA1F	16	20	153	196	10	19	81	148

name and a scan slice number are used to name each image. Also, the corresponding labels have been obtained with the base name. Class balance in the cross-validation approach was guaranteed by distributing cases with class 0s and class 1s equally among folds after counting the total number of label 0s and label 1s, respectively. Then, to stop data leaks during model training, all slices with identical base names were placed in the same folds. Label encoding was used to convert categorical variables into numerical codes, ensuring uniform labeling across classes and simplifying numerical representation. The step-by-step process utilized to accomplish this is shown in detail in figure 3.5.

The preprocessing step was essential because it resized and normalized images to minimize undesired biases while ensuring they were prepared to be used by the model for further processing. Normalization values and pixel sizes were adjusted and applied based on the requirements of the pre-trained architecture.

All the images were resized to dimensions of 224x224 pixels to align with the input

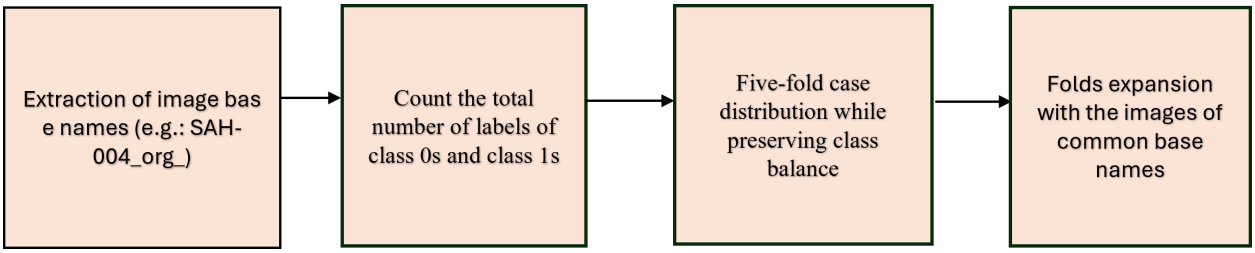

Figure 3.5: A step-by-step process for evenly distributing cases while controlling for class imbalance and preventing data leakage

requirements of the selected architectures, followed by normalization for pixel adjustment, aimed at boosting the model's efficacy. Most image pixel values typically range between 0 to 255, but it's ideal to scale them between 0 and 1 to enhance model construction efficiency by reducing the computational complexity of model training. Besides, complementary data from a single slice was integrated to form a 3-channel image. Initially, the analysis relied on all three channels representing the original CT image. However, upon further evaluation, it was observed that incorporating distinct layers of information into the three channels improved model performance in predicting outcomes, providing additional insights. Thus i) the original raw CT image, ii) the segmented CT labeled image (including cerebrospinal fluid, gray matter, white matter, and blood regions), and iii) only blood markings of the slices were extracted as the different inputs to three channels, respectively, to construct the stacked input image as shown in an example in figure 3.6 for deep learning-based prediction models.

3.2.3 Proposed Deep Learning Pipeline

The study explored the efficacy of deep learning architecture in extracting relevant imaging biomarkers. After testing multiple architectures (i.e. Dense-121, ResNet-50, Vgg16, Inception), Dense121 architecture and Vgg16 architecture were selected as the pertaining architectures of the proposed model based on the initial performance and training speed evaluation. Several studies have shown the effectiveness of these architectures in achieving successful classification and prediction performance. Both architectures served as pre-training models

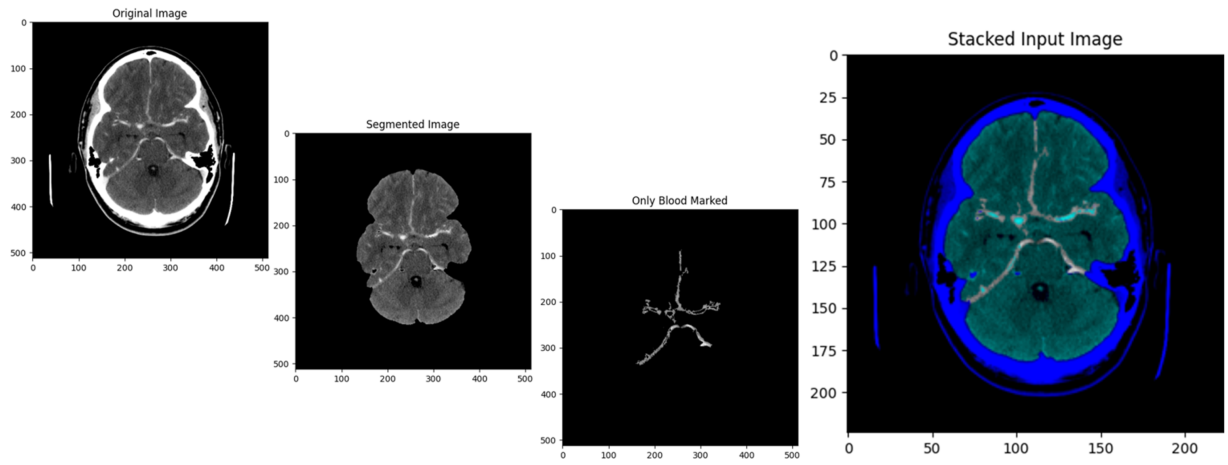

Figure 3.6: An example illustrating the construction of an input (original) image for feature extraction, achieved by combining complementary information from a CT slice

for feature extraction, followed by the incorporation of layers and attention mechanisms tailored to our specific task. Data augmentation was implemented in each fold to address the small dataset and mitigate overfitting concerns. The model's performance was subsequently evaluated using selected metrics to demonstrate the effectiveness of the prediction model.

3.2.3.1 Data Augmentation

Data augmentation is crucial to generating synthetic samples and constructing a diverse dataset—particularly when the original dataset is comparatively small. This process involves creating heuristic image data and increasing the number of training samples through image modifications while preserving the semantic information. Common strategies applied for data augmentation in model training include random cropping, color space augmentations, and horizontal flipping. These techniques help mitigate bias in the image data caused by similarity to the underlying images [123]. Consequently, overfitting is prevented, enabling the model to learn a broader set of features.

Every image in this study was subjected to unique augmentation in each of the five folds, incrementally enhancing each training fold with each iteration. The eight augmentations applied to each image included 45 degrees of rotation, 0.2 degrees of shear, 0.2 degrees of

vertical and horizontal shift, horizontal flipping, 0.2 degrees of zoom, and constant fill mode. These parameters were meticulously chosen to enrich the training dataset to encompass a diverse range of variations, such as rotations, shearing, shifting, flipping, and zooming. This strategy aimed to improve the model's robustness and its ability to generalize effectively across various input image variations.

3.2.3.2 Transfer Learning

Achieving the optimized performance with the available CT slices necessitated a large training set to improve the performance of predictive models. Transfer learning (TL) is a well-established strategy to address such issues. The TL technique involves pre-training a model on a large-scale labeled dataset and subsequently utilizing the learned weights for similar tasks within the same architectural framework. By leveraging the pre-trained model as an initial point in the training process for the target task, TL eliminates the need to commence model training from scratch with random weight initializations. In our study, we employ Dense121 and VGG16 architectures, both pre-trained on the ImageNet dataset comprising over 14 million images.

- **Dense121:** The DenseNet architecture, pre-trained on large datasets, proves to be highly effective when applied to small datasets. Its dense connections allow it to capture relevant insights, bypass traditional layers, and promote feature reuse. Figure 3.7 illustrates the connectivity in the DenseNet architecture, where layers are linked by dense blocks. Each layer generates a feature map that passes information to subsequent layers by integrating inputs from all preceding layers [107]. This connectivity enhances information flow, reducing overfitting. This approach is crucial for small datasets as it prevents the model from learning specific patterns, helping it to identify more generalizable features and maximizing the information extracted from each data point.

  Among the several DenseNet architectures available, such as Dense160, Dense121, and Dense201, we concentrated on the Dense121 architecture in this study. Dense121

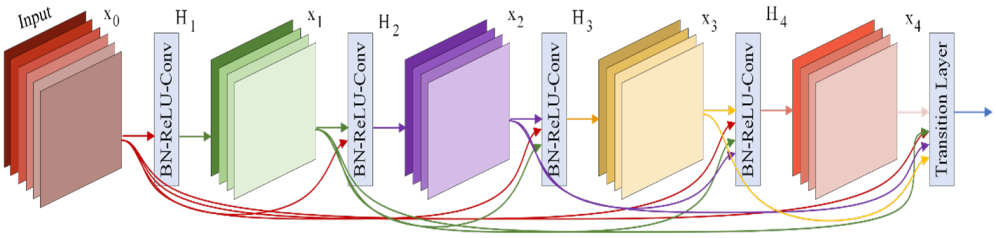

Figure 3.7: Dense Net Architecture: dense blocks are characterized by each layer receiving feature maps from all preceding layers as input[6]

comprises three transitional layers that control the feature map's size through down-sampling and four dense blocks intended to improve information flow.

Table 3.3 illustrates the architecture of the proposed model, which consists of a DenseNet 121 backbone serving as a feature extractor and several densely linked layers for classification. 701,313 of the 7,742,785 parameters in it are trainable. After the input was processed through the DenseNet121 layers, 7x7x1024 feature maps were produced. The next step was to use global average pooling to get a 1024-dimensional feature vector. This vector was passed through a series of dense layers (512,256, 128, 64) with progressively fewer neurons in each layer. Batch normalization and dropout were applied for regularization in each layer. A binary classification task was indicated by the final output generated by a single neuron.

- **VGG16:** VGG16 remains a potent technique for image classification, even with its advanced age, particularly when data is scarce. Praised for its success in the 2014 ImageNet competition, VGG16 continues to be the approach of choice for various image classification applications. Its straightforward and consistent architecture, featuring layered convolutional filters, excels at extracting features from limited datasets. By leveraging transfer learning and pre-training on the extensive ImageNet dataset, VGG16 can adapt its general knowledge to specific tasks. Fine-tuning the pre-trained weights for particular applications allows VGG16 to identify patterns even in scenarios with sparse data, making it a reliable solution for image classification challenges with

Table 3.3: Proposed model architecture developed on pretrained Dense121

Layer (type)	Output Shape	# of Parameters
densenet121(Functional)	(None,7,7,1024)	7037504
globalaveragepooling2D	(None,1024)	0
batch normalization1	(None, 1024)	4096
dropout1 (Dropout)	(None, 1024)	0
dense1 (Dense)	(None, 512)	524800
batch normalization2	(None, 512)	2048
dropout2 (Dropout)	(None, 512)	0
dense2 (Dense)	(None, 256)	131328
batch normalization3	(None, 256)	1024
dropout3 (Dropout)	(None, 256)	0
dense3 (Dense)	(None, 128)	32896
batch normalization4	(None, 128)	512
dropout4 (Dropout)	(None, 128)	0
dense4 (Dense)	(None, 64)	8256
batch normalization5	(None, 64)	256
dropout5 (Dropout)	(None, 64)	0
prediction layer (Dense)	(None, 1)	65
Total params: $7,742,785$		
Trainable params: $701,313$		
Non-trainable params: $7,041,472$		

Table 3.4: Pretrained VGG16 transfer learning architecture

Layer (type)	Output shape	Param #
Vgg16 (Functional)	(None, 7, 7, 512)	14714688
flatten1 (Flatten)	(None, 25088)	0
dense1 (Dense)	(None, 128)	3211392
dropout1(Dropout)	(None, 128)	0
Prediction layer (Dense)	(None, 1)	129
Total params: 17,926,209		
Trainable params: 3,211,521		
Non-trainable params: 14,714,688		

limited resources.

The model employed a VGG16 backbone for feature extraction, followed by a fully connected decision-making head, encompassing a total of 14,714,688 trainable parameters. The input underwent processing through VGG16 layers, producing 7x7x512 feature maps. These maps were then flattened into a 25,088-dimensional feature vector. This vector traversed a dense layer with 128 neurons for additional processing, followed by a dropout layer for regularization. Finally, a single-neuron dense layer generated the model's output, indicative of a binary classification task.

3.2.3.3 Transfer learning model integration with attention mechanism

In deep learning, attention models provide numerous advantages as discussed in section 2.4, particularly regarding image classification with sparse data. Especially when the dataset is small, attention techniques are essential for improving the model's focus on pertinent elements and maximizing performance. A notable development in the field of deep learning for computer vision is the Convolutional Block Attention Module (CBAM). CBAM presents an attention mechanism that improves convolutional neural networks (CNNs) by dynamically emphasizing relevant spatial and channel-wise information inside input feature maps [182].

Through implementing this, the network is more competent in extracting important traits and suppressing less significant ones. As demonstrated in figure 3.8, CBAM combines channel and spatial attention mechanisms to provide a thorough focus on substantial features.

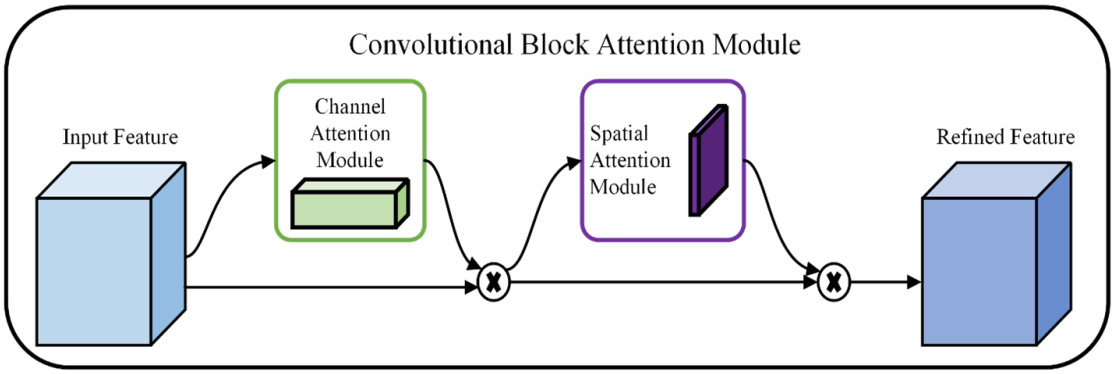

Figure 3.8: Convolution Block Attention Mechanism [7]

So in summary, the model incorporated fully connected layers and attention methods for prediction by classification, and it used Dense121 and Vgg16 backbones for feature extraction. The procedure proceeded in three distinct phases: Feature Extraction (using Dense121 and Vgg16 to extract features); Global Average Pooling (which generates a combined feature vector); and Attention Mechanisms (which involved channel and spatial attention modules to sequentially refine the feature vector to highlight informative channels and spatial regions). Then, the fine-tuned feature vector passed through Dense Layers consisting of successively fewer neurons. The resulting fine-tuned feature vector followed Dense Layers with progressively fewer neurons, each with batch normalization and dropout applied for regularization. In the last step, a prediction layer with a single neuron produces an output that resembles a binary classification problem for the model. Through careful feature extraction, careful refinement, and well-informed classification, the model was made possible by this all-encompassing methodology. Multiple performance indicators are employed to assess the prediction model in the end. The complete workflow of the complete proposed model architecture, featuring all the significant steps, is illustrated in figure 3.9.

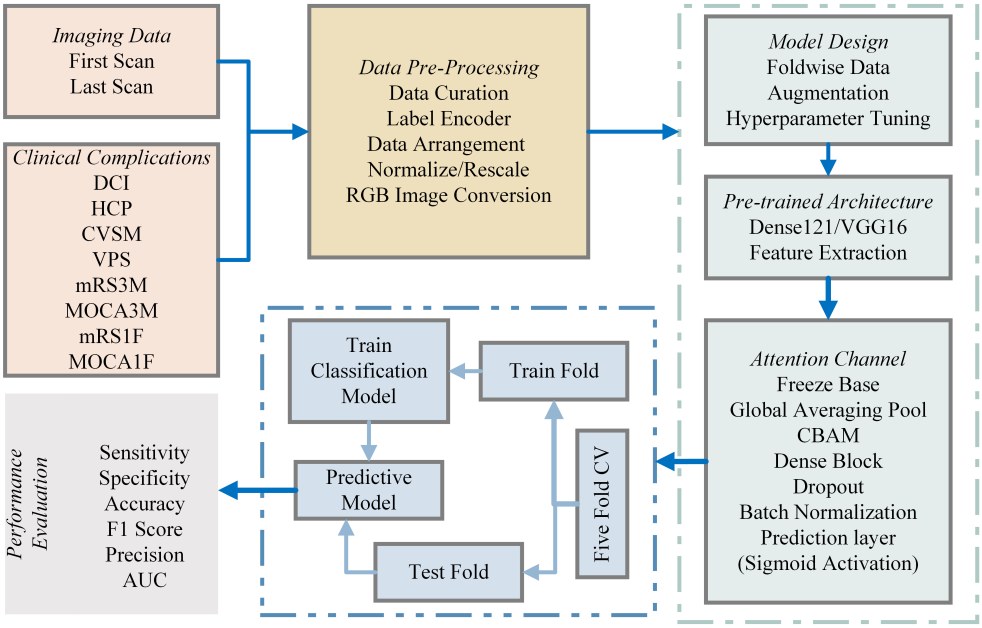

Figure 3.9: Flowchart of the proposed model using transfer learning model with attention (CBAM) mechanism

3.3 Model Implementation and Evaluation

3.3.1 Hyperparameter Tuning

All the experimentation was conducted using Python 3.9 within Jupyter Notebook, implementing TensorFlow/Keras flow libraries. The training sessions were executed on a local machine equipped with a 12th Gen Intel(R) Core (TM) i9-12900H processor, 32GB RAM, and an NVIDIA GeForce RTX 3080 (8GB) graphics card. Furthermore, Google Colab Pro was leveraged to access a more powerful GPU. The optimization process employed the Adam optimizer for both Dense121-CBAM and VGG16-CBAM architectures. A learning rate of 1e-3 and 1e-4 was set for the respective models, with a momentum value of 0.9 and no decay. The training spanned 150 epochs, with early stopping implemented to address potential overfitting. L1/L2 regularization was applied with rates of 1e-5 and 1e-4 for Dense121-CBAM and VGG16-CBAM, respectively. The model processed images in batches of 64, utilizing the he(normal) initializer for weight initialization with a ReLU-based activation function. The

'balanced' argument in compute-class-weight was used to determine class weights to control class imbalance while assigning weights to each class based on its frequency in the training data, helping the model treat each class fairly during training. In the final prediction layer, the sigmoid activation function was employed to predict binary classes while weight was initialized with glorot (normal) initializer. This configuration effectively balanced pre-trained features, introduced new layers, and employed regularization techniques, optimizing performance within the specified 150 epochs.

3.3.2 Performance Evaluation Metrics

Two statistical data analyses were employed methods to assess and compare the predictive performances of the aforementioned architectures. Initially, a ROC type data analysis approach was utilized, employing the AUC constructed from the classification score as an evaluation metric. Subsequently, the model's performance was assessed by applying an operating threshold (T = 0.5) during each training fold on the classification scores, dividing all testing cases into two classes ('Class-1' for scores $0.5 \leq$ and 'Class-2' for scores > 0.5). Several confusion matrices from the classification results were generated, which are utilized to calculate various performance indices for class classification, including classification accuracy (ACC), sensitivity, specificity, precision, recall, and F1. To obtain an overall evaluation, these performance values across the 5-fold were averaged, providing a comprehensive assessment of the model's performance on the entire dataset, as reported in the next section.

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.2}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.3}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.4}$$

$$F1_{measure} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.5}$$

3.4 Results

Figure 3.10 depicts the influence of incorporating CBAM into pre-trained architectures on model performance, specifically examining sensitivity, F1-score, and AUC values for DCI, HCP, and mRS at 3 months across admission and discharge scans. For DCI, sensitivity ranges from 0.49 to 0.8 , with the VGG-16-CBAM architecture achieving the highest value of 0.88 (95% CI, 0.782, 0.988) compared to 0.63 (95% CI,0.518, 0.7516) for the base VGG-16 using the admission scan. Dense121-CBAM performed supremely well with a sensitivity of 0.86 (95% CI, 0.755, 0.965), while the base Dense-121 architecture gave 0.65 (95% CI, 0.517, 0.7896). Similarly, higher sensitivity values for HCP and mRS at 3 months are accomplished by VGG-16 with CBAM, with values of 0.97 (95% CI 0.94, 1.002), and 0.91 (95% CI 0.7748, 1.037), respectively.

Regarding the F1-score, both the Dene121 and VGG16 CBAM architecture performed well with DCI, reaching a peak value of 0.81 (95% CI, 0.775, 0.855) and 0.82 (95% CI, 0.726, 0.920) respectively. In contrast, Dense121 without CBAM reports the lowest F1-score for DCI at 0.48 (95% CI, 0.4173, 0.5521). For mRS at 3 months, the highest F1-score is obtained from VGG16 with CBAM 0.69 (95% CI, 0.638, 0.733), while default VGG16 gave an F1-score of 0.57 (95% 0.507, 0.638). AUC scores range from 0.4360 ± 0.09746 to 0.7820 ± 0.0632 with the lowest reported from Dense121 and the highest attained from VGG16 with CBAM architecture for DCI.

Figure 3.11 and figure 3.12 showcase the performance improvement of ROC curves for both short and long-term clinical measures. The plots present ROC curves across five-fold cross-validation, featuring mean ROC curves from both Dense-121 (with and without CBAM) and VGG16 (with and without CBAM) architectures. The addition of the attention mechanism to the pre-trained architecture leads to slight improvements in accuracy and precision, but specificity metrics do not show the anticipated enhancement.

The model also assessed the differences in short- and long-term clinical measures between the admission and discharge scans. In figure 3.13 a comparison of performance metrics, specifically sensitivity, and accuracy, for clinical measures is presented between the admission

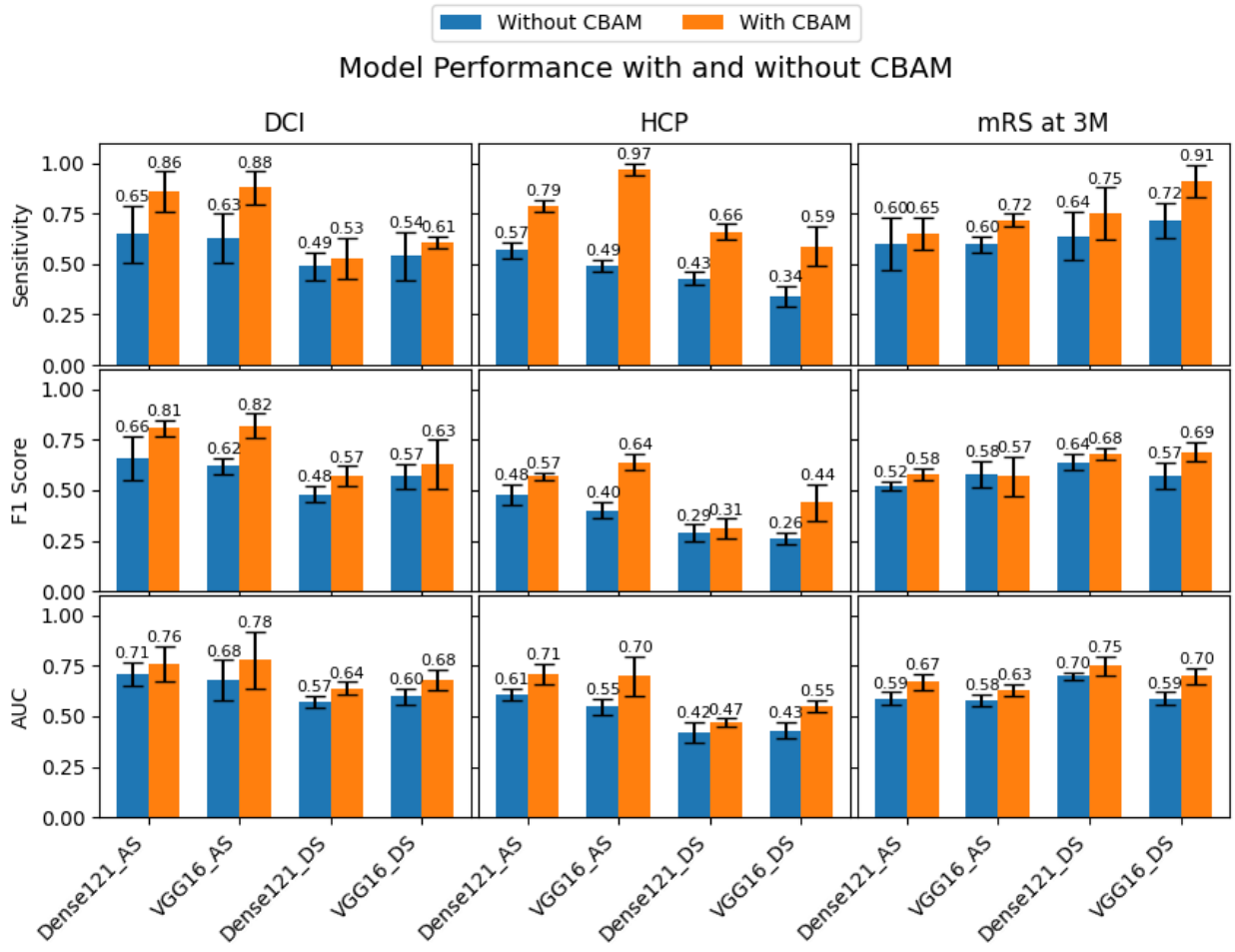

Figure 3.10: Performance comparison of models with and without CBAM; AS and DS denote admission scan and discharge scan, respectively.

and discharge scans. It is noteworthy that DCI and HCP are more accurately predicted using the first CT scan. For short-term measures, sensitivity reported from the admission scans is 0.86 (95% CI, 0.755, 0.965) and 0.88 (95% CI, 0.782, 0.988) for Dense121-CBAM and VGG16-CBAM models, respectively, whereas values of 0.53 (95% CI, 0.356, 0.711) and 0.61 are reported from the discharge scans for the same architectures. Similarly, the accuracy metrics of the models exhibit a range between 0.43 and 0.74, with the smaller value reported from the discharge scan and the higher value from the first scan. Conversely, for the long-term measure of mRS at 3 months, the last scan reported a classification accuracy for VGG16-

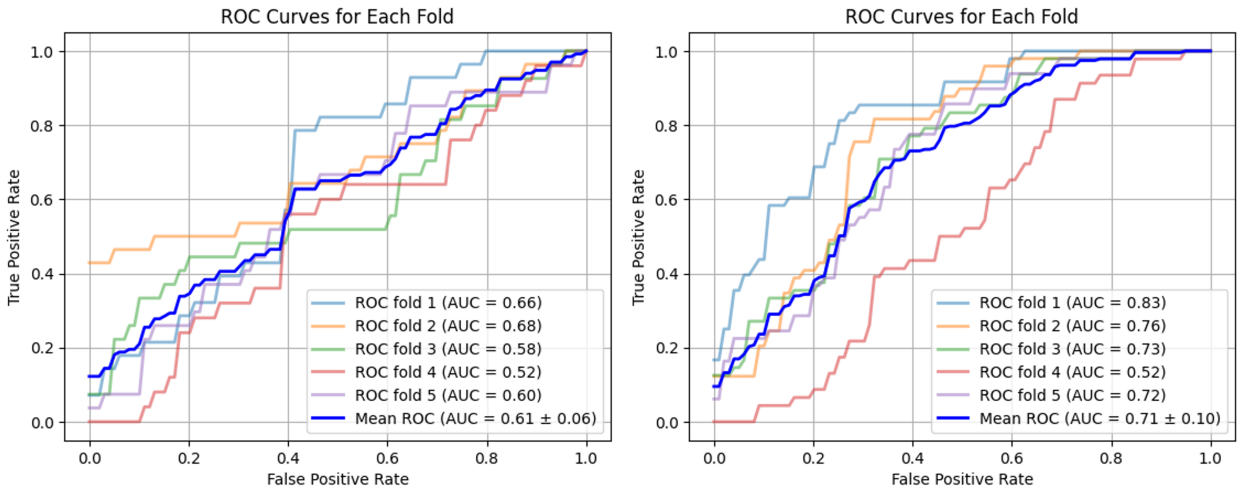

Figure 3.11: A comparison of ROC curves for short-term clinical measure HCP between the architectures. (left) represents the plot from Dense121 architecture, and (right) represents the plot from Dense121-CBAM architecture

CBAM is 0.60 (95% CI, 0.421, 0.733), compared to 0.51 (95% CI, 0.4285, 0.610) from the admission CT scan. The sensitivity of long-term mRS was also observed to be high at 0.91 (95% CI 0.7748, 1.037) with the discharge scan as opposed to 0.72 (95% CI, 0.637, 0.811) from the admission scan. This observation suggests the influence of treatment during the hospitalization of patients on long-term outcomes.

Table 3.5 illustrates more comparisons between the first and last scans using additional performance evaluation metrics, including AUC, precision, F1 score, and specificity for all three clinical measures. Notably, the specificity performance of DCI yielded a different result, showing a lower value of 0.44 (95% CI, 0.394,0,484) from VGG16-CBAM architecture from the admission scan, whereas the score with the discharge scan was 0.62 (95% CI, 0.3398, 0.9126) for the same model. However, the AUC value, precision, and F1 score demonstrated better performance with the admission scan for short-term DCI and HCP, while mRS-3M gave a better result for those evaluations from the discharge scan. The highest recorded F1 score value for DCI was 0.81 (95% CI, 0.775, 0.855), which came from assessing the admission scans using Dense121-CBAM architecture, where the value from the discharge scan from the

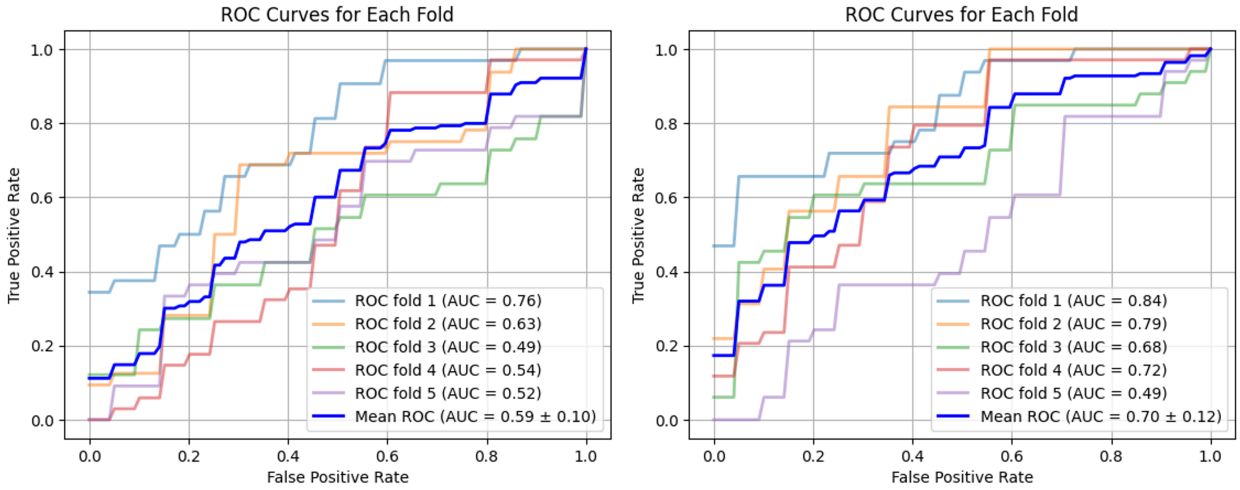

Figure 3.12: A comparison of ROC curves for long-term clinical measure mRS at 3 months between the architectures. (left) represents the plot from VGG16 architecture, where (right) represents the plot from VGG16 with CBAM architecture

same model was 0.77 (95% CI, 0.712,0.913). Similarly, the AUC score for mRS-3M reported the highest value of 0.70±0.12 from the discharge scan using the Dense121-CBAM model, where the lowest value observed was 0.63 ± 0.1 from the admission scan.

Table 3.5: Performance metrics from Dense121-CBAM architecture and VGG16-CBAM architecture

Model	Scan Type	Clinical Measure	AUC	Precision	Specificity	F1-Score
Dense121-CBAM	First	DCI	0.76 ± 0.21	0.78	0.52	0.81
		HCP	0.71 ± 0.10	0.51	0.63	0.57
		mRS3M	0.59 ± 0.12	0.61	0.61	0.58
	Last	DCI	0.57 ± 0.19	0.65	0.42	0.77
		HCP	0.42 ± 0.11	0.31	0.26	0.42
		mRS3M	0.75 ± 0.18	0.56	0.64	0.63
VGG16-CBAM	First	DCI	0.78 ± 0.14	0.67	0.44	0.76
		HCP	0.70 ± 0.08	0.53	0.48	0.65
		mRS3M	0.63 ± 0.1	0.63	0.74	0.65
	Last	DCI	0.68 ± 0.15	0.60	0.62	0.62
		HCP	0.42 ± 0.15	0.47	0.33	0.63
		mRS3M	0.70 ± 0.12	0.69	0.81	0.65

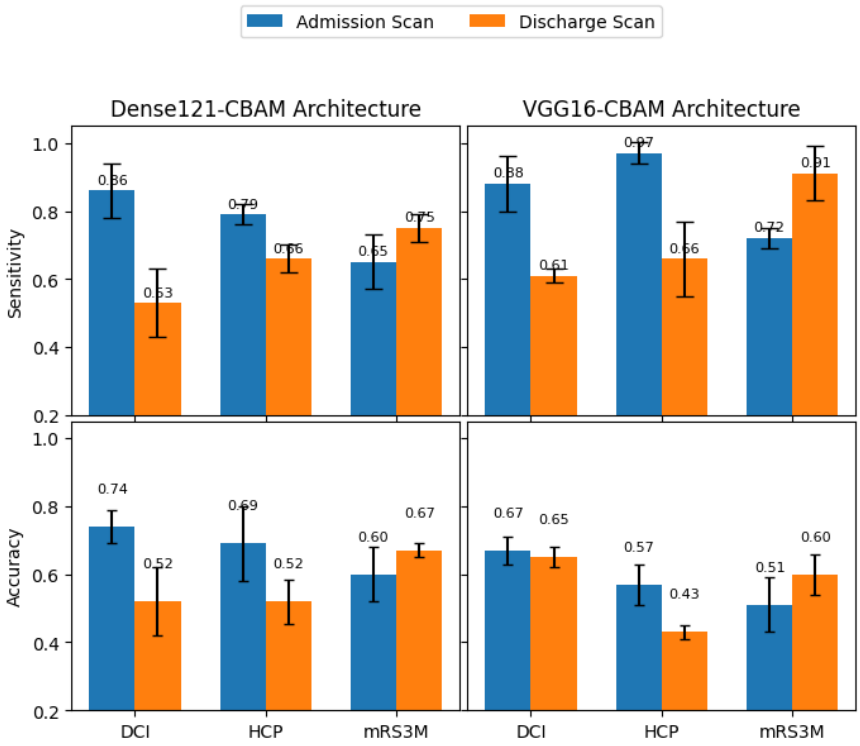

Figure 3.13: A side-by-side comparison of admission scan and discharge scan between performance metrics sensitivity and accuracy from two architectures

3.5 Discussion and Conclusion

This study highlighted the significant capability of deep-learning-based classification models in predicting both short-term and long-term prognosis for patients after aSAH. Three clinical outcomes of aSAH were explored, which are currently evaluated subjectively and qualitatively by neuroradiologists in clinical settings. However, this poses a challenging task with the potential for high inter-rater variability. Several clinical biomarkers have been reported to investigate the impact of aSAH on the development of EBI and, consequently, on DCI. A recent report demonstrated the potential of imaging markers in clinical research, aiding clinicians in constructing prediction models, a domain not widely explored for predicting the effects of aSAH. To further harness the potential of imaging markers while considering the established risk factor of a dense cisternal clot on a CT scan for DCI after SAH, an imaging

biomarker using a deep-learning model was proposed in this retrospective study.

The model considered the initial and final CT scans, along with information on the clinical outcomes of the same cases, for performance assessment. During the model development, initially, an effort was made to identify the most suitable foundational architectures based on the data's size and type, and Dense121 and VGG16 were selected accordingly to serve as pre-trained models for feature extraction. Subsequently, to enhance focused learning, the CBAM attention model was incorporated into the pre-trained architectures, resulting in a significant improvement in the model's performance metrics.

The sensitivity metric, gauging the model's ability to identify patients with a positive outcome after aSAH accurately, yielded high results of 0.86 (95% CI, 0.755, 0.965) for Dense121-CBAM architecture and 0.88 (95% CI, 0.782, 0.988) for VGG16-CBAM architecture for predicting DCI. These results marked an improvement from the default architectures, which had sensitivities of 0.65 (95%, 0.485,0.822) and 0.63 (95 % CI, 0.220, 1.052), respectively, demonstrating the positive impact of adding CBAM to the model. Similarly, sensitivity for HCP and mRS at 3 months significantly improved with the introduction of the CBAM model to the base architectures. This heightened sensitivity suggests that the model can effectively identify patients at a high risk of poor outcomes, as missing these patients could delay interventions and worsen their prognosis. The study also reported high AUC scores for all three clinical measures when utilizing attention models in both architectures. The AUC metric, signifying the model's capacity to discriminate between positive and negative classes, exhibited scores of 0.6840 ± 0.1478 and 0.6960 ± 0.0838 for DCI and HCP, respectively, with the VGG16 CBAM architecture, as opposed to 0.6360 ± 0.082 and 0.5040 ± 0.2194 from the base architecture. Furthermore, the AUC score reported for mRS at 3 months ranged between 0.7320 ± 0.1859 and 0.7540 ± 0.1783 for both architectures, surpassing the range for the default architectures of 0.6000 ± 0.1863 and 0.6440 ± 0.1609. These robust AUC values underscored the importance of the CBAM architecture and suggested that the model excels in effectively distinguishing individuals with the condition from those without. The research also documented a substantial F1 score of DCI equal to 0.81 (95% CI, 0.726, 0.920), affirming

a balance between accuracy and thorough identification of high-risk patients by mitigating both false positive and false negative cases.

Furthermore, the study presented an insightful observation emphasizing the importance of the scan period in predicting prognosis. It was noted that the initial CT scan taken on the admission day offers a more precise prediction for short-term clinical measures, including DCI and HCP. Conversely, the CT scan conducted 10-14 days after admission proves more accurate for the long-term measure, mRS at 3 months. As illustrated in figure 3.13 and detailed in table 3.5, processing the images from the admission scan through our prediction model yields higher values for sensitivity, accuracy, AUC score, F1 score, and precision for DCI and HCP. With the VGG16-CBAM model on the first scan, the score for DCI sensitivity was 0.88 (95% CI, 0.782, 0.988), compared to 0.67(0.485, 0.80) from the last scan using the same model. In the case of HCP, the accuracy metrics for the first scan with Dense121-CBAM architecture were 0.69 (95 % CI, 0.586,0.794), significantly surpassing the 0.52 (95% CI, 0.442, 0.605) accuracy from the last scan. This suggests that using this image marker from the initial scan has the potential to precisely predict short-term clinical outcomes, which can assist clinicians in providing valuable insights for fast decision-making to control EBI. On the contrary, when assessing long-term measures, we observed that the sensitivity of mRS at 3 months from the first scan was 0.72 (95% CI, 0.5881, 0.8555) , showing a substantial increase to 0.91 (95% CI, 0.795, 1.037) while training with the last scans. This observation emphasizes the impact of patients' treatment on long-term clinical outcomes, offering clinicians valuable insights for formulating optimal tailored recovery plans upon patients' discharge from the hospital.

While positive aspects have been identified, this study contends with several limitations. The holistic nature of deep architectures can sometimes hinder clinical acceptability. However, integrating domain knowledge with deep features could create a more advanced imaging biomarker, improving its clinical relevance and acceptance. Moreover, the study focused on only three clinical outcomes. Expanding the assessment to encompass a broader range of prognoses would contribute to establishing a more robust model for improved clinical pre-

dictions in future applications.

In conclusion, despite certain limitations, the study successfully demonstrated the potential of a prediction model for aSAH prognosis, leveraging blood clot information from CT images to assess three clinical measures. The observed higher sensitivity, AUC score, and F1 score indicate the model's strong capability to identify patients with potential positive outcomes, enabling quicker interventions by medical professionals. Additionally, the study underscored the importance of the CT scan's time period in evaluating both short-term and long-term outcomes. Building on the foundation laid in this research, future efforts can focus on refining and validating these innovative quantitative image markers through prospective clinical studies. And towards the effort to avail a more clinically acceptable prediction model, my next study focuses on integrating radiomic features with deep features which will add more domain knowledge to the model,

CHAPTER 4

RaDIS-aSAH: the Radiomic and Deep Learning Integration System to Predict Clinical Events and Outcomes After Aneurysmal Subarachnoid Hemorrhage

4.1 Introduction

Many researchers have embraced a hybrid approach, integrating both Machine Learning (ML) and Deep Learning (DL) techniques, as evidenced by numerous studies exploring this integration [30, 114, 183]. This approach capitalizes on the complementary strengths of both AI models. Several studies reported the successful execution of several hybrid approaches in the context of different brain disease prediction [184, 125, 185, 186]. In my second study, I propose the integration of the features from the radiomic and DL model, RaDIS, to establish a hybrid ensemble model for predicting short-term clinical events and future outcomes in aSAH patients. Initially, in our pilot investigation, we engineered a CAD system with a primary emphasis on generating radiographic image markers associated with sulcal volume to address complications following aSAH [187]. Subsequently, the scope of the study

was expanded to predict aSAH complications by employing DL frameworks, extracting intricate deep features from slices with marked blood region [188]. An imaging biomarker was devised centered on blood volume within ventricular and cisternal spaces, utilizing sophisticated deep-learning architectures to measure the varied hemorrhagic features apparent in CT scans. These studies have shown promising results in highlighting the feasibility of developing radiomic image markers to predict a subset of clinical complications in aSAH patients. However, no prior research has been conducted to explore hybrid ensemble approaches that combine radiomics, machine learning, and deep learning frameworks on a wide range of short- and long-term complications in aSAH patients. Thus, in this following study to explore the efficiency of a hybrid approach, features from both studies were combined in various permutations to construct the ensemble-based model.

The objective of my second study was to develop a CAD system for aSAH patients using an ensemble model development that combines the strengths of radiomics and deep learning components. Subsequently, all three approaches—radiomics, deep learning, and hybrid—were examined and compared to determine the optimal prediction model for clinical outcomes in aSAH patients. The ultimate goal of all the studies was to construct an improved prediction model based on image biomarkers capable of categorizing the probability of patients experiencing various short-term and long-term clinical complications due to aSAH. To our knowledge, no prior studies have employed an ensemble technique for predicting outcomes in aSAH patients.

4.2 Flow diagram of the Proposed Ensemble-based CAD Scheme

The proposed computer-aided diagnosis (CAD) scheme ensembles models that combine radiomic features with deep features. Radiomic features focus on quantifying the volume of sulci across various brain regions, while deep features identify blood clots from hemorrhagic stroke in the subarachnoid space. This ensemble approach is particularly significant as it

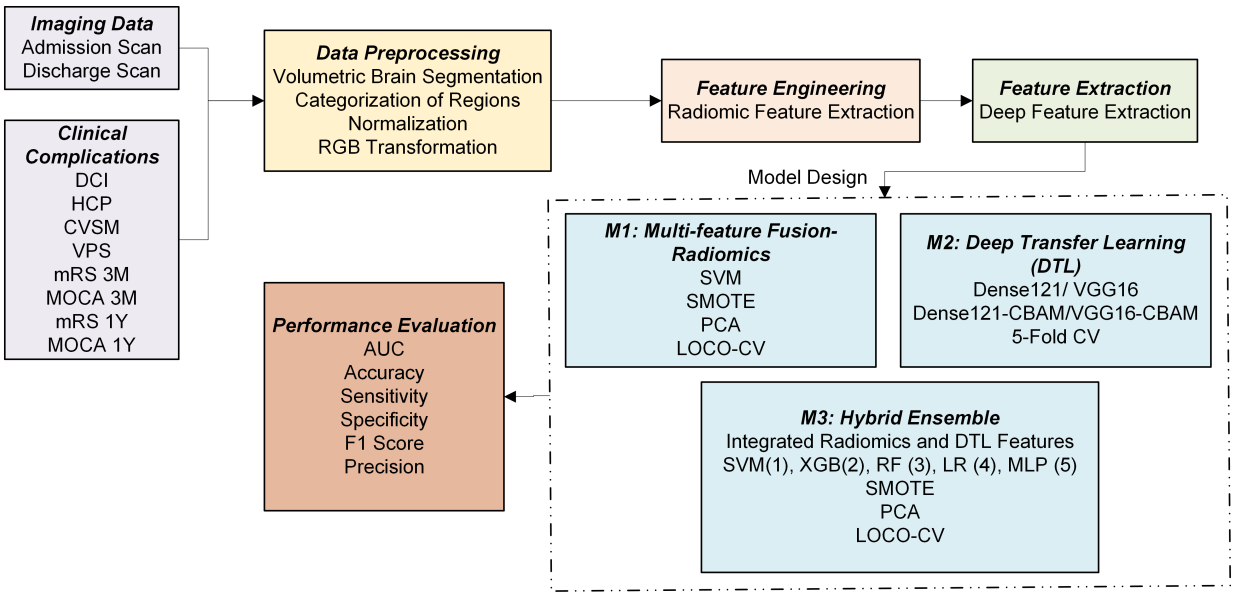

Figure 4.1: Flow diagram of proposed RaDIS-aSAH architecture

combines the complementary strengths of both models in predicting and aiding decision-making for patients with aSAH. The interpretability and robustness of the machine learning (ML) model in quantifying radiomic features, coupled with the intricate patterns captured by deep features, enhance the discriminative power of the prediction model. This aids clinicians in treatment planning for aSAH patients by providing comprehensive insights considering multiple factors. Figure 4.1 illustrates the flow diagram of the proposed ensemble approach, which consists of several stages. The data preprocessing phase of the CAD scheme begins with the segmentation of brain structures from raw CT scan data, followed by the categorization of distinct regions including white matter, gray matter, blood, cerebrospinal fluid, and sulci based on CT Hounsfield Unit values. Additionally, the blood regions within subarachnoid spaces were identified and stored as a segmentation mask, which served as a critical element of the model architecture. Each CT slice underwent normalization and transformation from grayscale to a 3-channel image to facilitate subsequent feature extraction.

Second, in the feature engineering phase, radiomic features were extracted by applying various techniques to the 3D CT scans, deriving 9 features based on the volume and

proportion of sulcal spaces compared to other brain regions. The third stage involves the automatic extraction of deep intrinsic patterns from CT slices, particularly focusing on the presence of blood region in the cisternal space, yielding 128 features through a deep transfer learning-based model. The fourth stage focused on building several models including multi-feature fusion radiomics, attention-based deep learning, and hybrid ensemble-based classification models. Subsequently, the performance of each prediction model was evaluated and compared using various metrics to assess its effectiveness pe within the CAD framework.

4.2.1 Radiomics Feature Engineering

Early reductions in selective sulcal volume have been confirmed as a potential quantitative indicator of edema in patients with aSAH. This offers the opportunity for earlier detection and continuous monitoring of brain swelling following aSAH. Thus, analyzing the volume of CT-labeled regions relative to sulci volume was a crucial stage in identifying and calculating radiographic image characteristics. Building on prior research, the analysis focused on image slices ranging from the level of the lateral ventricle to the top of the skull. The regions that are shown in the figure 4.2 identified as CSF by the threshold algorithm in the CAD scheme were recognized as sulcal regions, considering that CSF occupies the spaces within the sulci. Within this range, the proportion of sulci was quantified as a qualitative assessment to investigate their correlation with aSAH.

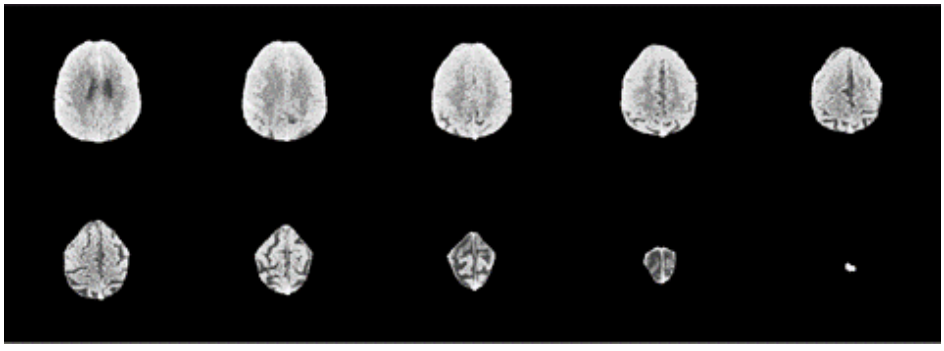

Figure 4.2: A representative CT scan showcasing the sulci region delineated within the specified analysis area

The first set of features was obtained as a volumetric representation of a single CT scan for each categorized region, including white matter (WM), gray matter (GM), blood, brain, and CSF/sulci. This was achieved by amalgamating image features across all selected CT slices. The volume for each labeled region within a single CT slice was computed using segmented voxel information (N_{Label}), voxel dimensions (V_L : voxel length and V_B: voxel breadth), and slice thickness (S_T). Table 4.1 outlines the equations utilized for calculating the features denoted as $Sulci_V$, WM_V, GM_V, $Blood_V$, and $Brain_V$ where"V" = volume from the sulci, white matter (WM), gray matter (GM), blood, and brain regions, respectively.

Table 4.1: Calculation of radiographic image features using 3D Volumes of each labeled region

Region	**Volume Per Slice**	**Volume of entire CT Scan**
Sulci	$Sulci_{V\|slice} = N_{Sulci} \times V_L \times V_B \times S_{T\|slice}$	$Sulci_V = \sum_{slice_{k=m}}^{n} Sulci_{V\|slice}$
WM	$WM_{V\|slice} = N_{WM} \times V_L \times V_B \times S_{T\|slice}$	$WM_V = \sum_{slice_{k=m}}^{n} WM_{V\|slice}$
GM	$GM_{V\|slice} = N_{GM} \times V_L \times V_B \times S_{T\|slice}$	$GM_V = \sum_{slice_{k=m}}^{n} GM_{V\|slice}$
Blood	$Blood_{V\|slice} = N_{Blood} \times V_L \times V_B \times S_{T\|slice}$	$Blood_V = \sum_{slice_{k=m}}^{n} Blood_{V\|slice}$
Brain	$Brain_{V\|slice} = N_{Brain} \times V_L \times V_B \times S_{T\|slice}$	$Brain_V = \sum_{slice_{k=m}}^{n} Brain_{V\|slice}$

Besides the five features mentioned above, another set of four additional features was computed representing the proportion volume of Sulci ($Sulci_V$) in a 3D CT scan with respect to volumes of other segmented regions of the brain (i.e.: WM_V, GM_V, $Blood_V$, and $Brain_V$). WM_{Sulci}, GM_{Sulci}, $Blood_{Sulci}$, and $Brain_{Sulci}$ are the newly generated features from the volumetric information using the following equations:

$$WM_{Sulci} = \frac{Sulci_V}{WM_V} \tag{4.1}$$

$$GM_{Sulci} = \frac{Sulci_V}{GM_V} \tag{4.2}$$

$$Blood_{Sulci} = \frac{Sulci_V}{Blood_V} \tag{4.3}$$

Table 4.2: Deep Learning model and extracted feature information

Deep Learning Architectures	# of Features
Dense121	128
Vgg16	128
Dense121-CBAM	128
VGG16-CBAM	128

$$Brain_{\text{Sulci}} = \frac{Sulci_{\text{V}}}{Brain_{\text{V}}} \tag{4.4}$$

As a result of the feature engineering process, we acquired a total of nine radiomic features from each 3D CT scan from this approach. The initial set of five features represents the volume of each labeled region of the segmented brain, while the subsequent set of four features provides information about the amount of sulci volume in the brain's 3D CT compared to other regions.

4.2.2 Attention-based Deep Transfer Learning for Feature Extraction

This study utilized a deep learning framework, as detailed in chapter 3, to automatically extract hierarchical features from images. To address the limitations of comparatively smaller dataset, transfer learning with two pre-trained CNN architectures: DenseNet-121 and VGG-16 were leveraged which were pre-trained on ImageNet's dataset of more than 14 million pictures classifying thousand distinct classes. To enhance these models' feature extraction capabilities, an attention mechanism known as the Convolutional Block Attention Module (CBAM) was incorporated into above deep transfer learning models. The performance power was enhanced by adding more weights to the relevant features of those architectures. set of 128 features and prediction scores were extracted from each deep architectures. These features were critical in improving the accuracy and robustness of our analysis. The table 4.2 summarizes the model names and the features extracted using this deep-learning approach.

4.2.3 Hybrid Ensemble CAD Modeling Pipeline

In this study, three distinct approaches were built and evaluated to predict short-term and long-term complications of aSAH patients. The three methodologies to build the models included multi-feature fusion-based radiomics *M1*, deep transfer learning methods with and without-attention mechanisms *M2*, and hybrid ensemble integrating both radiomic and deep transfer learning approaches *M3*; each contributing to enhanced diagnostic precision and reliability. Performance metrics were compared and assessed for the hybrid, deep learning, and radiomic techniques to identify the most efficient model for each outcome. Concerns like class imbalance in cases, redundancy and inconsistency in the feature pool, and optimal findings using a very small dataset were evaluated and carefully handled in all approaches.

In the radiomic method *M1* several advanced ML algorithms were utilized and meticulously analyzed the extracted features discussed in section 4.2.1 to discern patterns strongly associated with the presence of each complication. Principal Component Analysis (PCA) for feature reduction and synthetic minority oversampling technique (SMOTE) for class imbalance were included in the model settings. These were both integrated into a Leave-on-Case-Out (LOCO)-cross-validation (CV) to maximize training capabilities and remove case partition bias. Among all the ML models, SVM consistently presented a better result for the radiomic fusion approach for all eight clinical outcomes. Additional details about feature engineering and other machine-learning techniques are available in the article [187].

In *M2* method, both Dense 121 and VGG16 with and without CBAM independently explore the ability to extract relevant information and classify outlined aSAH prognoses. All trainable layers were frozen, retaining only the top layer for acquiring transfer learning features to train the prediction model with the available dataset. Resized, normalized, and stacked images were utilized for model training with pre-trained weights. As detailed in chapter 3, these deep transfer learning models were trained using the Adam Optimizer, the ReLU activation function, while the prediction layer was trained with the sigmoid activation function along with the binary cross entropy loss function to predict two classes of each outcome. Additionally, the class imbalance was addressed by adjusting class weights during the

5-fold cross-validation. This holistic approach enabled effective automatic feature extraction, attentive refinement, and informed classifications.

Finally, in *M3*, a hybrid model was created to provide an automated pipeline that combined radiomic feature engineering and deep transfer learning features seamlessly, utilizing the advantages of both approaches to create reliable ensemble models. The model ensured a patient-level data analysis by shifting the focus of deep learning features from individual slices to a broader patient-level perspective. Instead of processing features on a per-slice basis, the deep learning features were averaged at the patient level, significantly reducing variance and noise typically associated with more granular data levels. Once normalized to this level, the deep learning features were then integrated with radiomic features. This step ensured the data was aligned into a structured form suitable for subsequent modeling phases. The unified feature set then consisted of both averaged deep learning features and radiomic features, poised to enhance the model's overall predictive capability. The features were standardized using Z-score normalization. This scaling technique adjusted the features to have a mean of zero and a standard deviation of one, ensuring that all features contribute equally to the model without any single feature dominating due to its scale. This standardization step is crucial for machine learning algorithms as they are sensitive to the input data's scale.

All data then led to the exploration phase, where an in-depth analysis was conducted to uncover underlying patterns and potential biases. During this phase, key aspects such as class imbalance were addressed, which was vital for understanding the distribution of categories within the data and ensuring the model did not favor one class over others. Skewness was also considered in the data to detect any asymmetry in the distribution of variables that could affect model predictions. To address this, a logarithmic transformation was applied to correct for skewness, thereby normalizing the distribution of our data. Additionally, correlations were explored within the data using heatmaps, which provide a visual representation of the relationships between features and complications. This approach helped in identifying any strong correlations that might influence the behavior of the predictive models, guiding the feature selection efforts. Feature reduction is achieved using PCA to retain 95% of the

variance. This approach ensured that most of the data's informational value was preserved, while greatly reducing the number of features.

4.3 Performance Evaluation and Comparison:

The core objective was to create an optimized predictive model for aSAH patients to evaluate their prognosis. To achieve the target, the three methods were compared to find the best model for each outcome. *M1* derived nine radiomic features based on sulcal volume in the brain region. *M2* applied deep transfer learning with attention mechanisms, subdivided into four variations: the first two used pre-trained Dense121 and VGG16 models exclusively (e.g., Method 2.2 represented the VGG16-based prediction model within the transfer learning framework). The latter two incorporated CBAM-enhanced versions of these models (Dense121-CBAM and VGG16-CBAM), with Method 2.3 involving features from Dense121-CBAM. *M3* combined radiomic features from *M1* with deep features from *M2* in four sub-methods (3.1–3.4), each integrating *M1*'s radiomic features with the respective sub-methods of *M2*. For example, Method 3.3 involved a hybrid model that merged radiomic features *M1* with deep features from Dense121-CBAM *M2.3*. Table 4.3 summarizes the study approaches, each evaluated with five distinct ML models, to identify the optimal prediction model across 16 outcomes for aSAH patients.

In the study, we employed two statistical data analysis methods to evaluate and compare the predictive capabilities of the aforementioned model approach. Initially, we evaluated the model's performance by applying an operating threshold (T = 0.5) to the classification scores during each training fold, dividing all testing cases into two classes ('Class-1' for scores $0.5 \leq$ and 'Class-2' for scores > 0.5). We generated multiple confusion matrices from the classification results, which were used to compute various class classification performance metrics, including classification accuracy, sensitivity, specificity, precision, recall, and F1 score. We then used ROC analysis to construct the area under the AUC from the classification scores as an evaluation metric. This comprehensive assessment will guide us in selecting the most effective predictive model among all the approaches discussed in Table

Table 4.3: Summary of approaches to the prediction model and associated feature information

Approach	Features
1	Radiomic Features (RF)
2.1	Dense-121
2.2	VGG16
2.3	Dense121-CBAM
2.4	VGG16-CBAM
3.1.x	RF+Dense121
3.2.x	RF+VGG16
3.3.x	RF+Dense121-CBAM
3.4.x	RF+VGG16-CBAM
3.5.x	RF+Dense-121 + VGG16+ Dense-121-CBAM + VGG16-CBAM

Note: *x indicates the ML algorithm models, where x corresponds to specific models: x = 1 represents SVM, x = 2 denotes XGB, x = 3 signifies RF, x = 4 stands for LR, and x = 5 refers to MLP.

based on overall performance and robustness.

4.4 Results

Figure 4.3 and 4.4 depict a comparative analysis of accuracy metrics across radiomic, deep learning, and hybrid approaches for short-term and long-term outcomes, considering both initial and final scans. Among 45 simulation results derived from various hybrid approach models, we identified the top two performance models for each outcome. In figure 4.3, the bar charts show accuracy metrics for DCI ranging from 0.63 to 0.83, with model 3.3.1 achieving the highest score of 0.83. For HCP, approach 3.5.2 reaches an accuracy of 0.9, within the range of 0.52 to 0.9, where *M1* and *M2* have shown the highest value as 0.72 and 0.61, respectively, highlighting the hybrid approach, *M3*, as the optimal predictive model.

Similarly, VPS ranges from 0.86 to 0.54, with approach 3.5.5 attaining the highest accuracy of 0.86. For CVSM, accuracy spans 0.61 to 0.78, with the highest value of 0.78 achieved by *M1*, closely matched by approach 3.3.2 of the hybrid model at 0.76. This indicates that *M3* approach provides superior or comparable predictive performance across all short-term prognostic metrics.

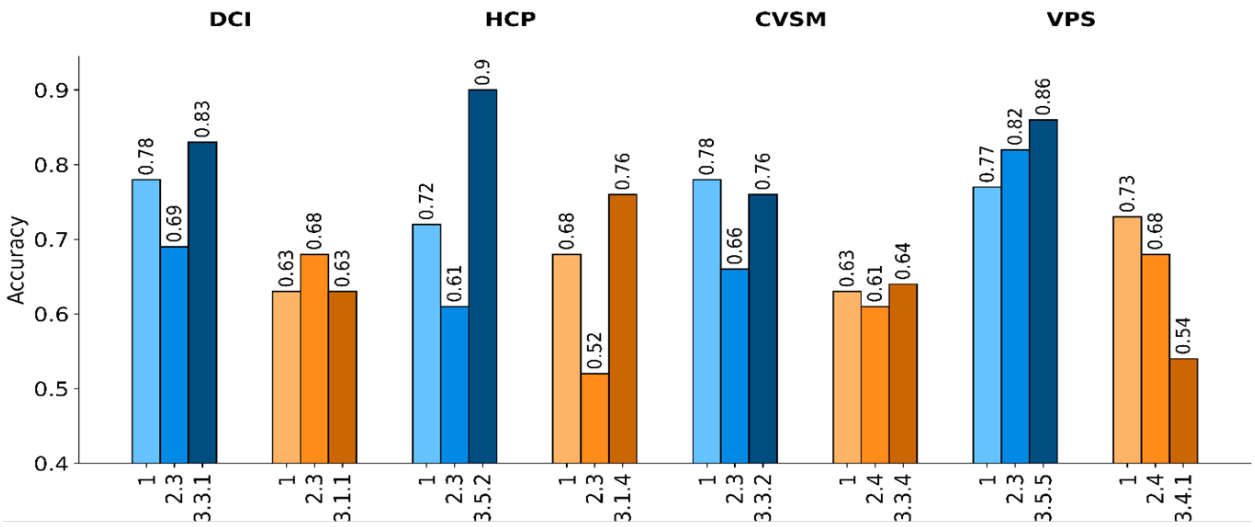

Figure 4.3: Comparison of accuracy metrics among radiomic, deep learning, and hybrid approaches for (top) short-term outcomes and (bottom) long-term outcomes

Assessing the long-term outcomes from figure 4.4, we observe that the mRS3M scores ranged from 0.63 to 0.84, with the highest value achieved by approach 3.5.1. The score for *M1* was 0.8, while the maximum score for *M2* (2.4) was 0.71, both of which were surpassed by the hybrid approach, *M3*. Similarly, for mRS1F, the highest accuracy was reported for 3.5.1, with scores ranging between 0.62 and 0.74.

Another long-term outcome, MOCA3M, also displayed its highest value from approach 3.5.1, with accuracy values ranging between 0.65 to 0.94. The second highest value has also been reported from approach 3.5.2, whereas approaches 1 and 2 have values of 0.78 and 0.77 respectively. For MOCA1F, 3.5.2 reportedly had the highest value of 0.85, surpassing the maximum value of 0.79 and 0.7 reported by approaches 1 and 2, respectively. These comparisons underscore the enhancement of the model when the hybrid approach is adopted.

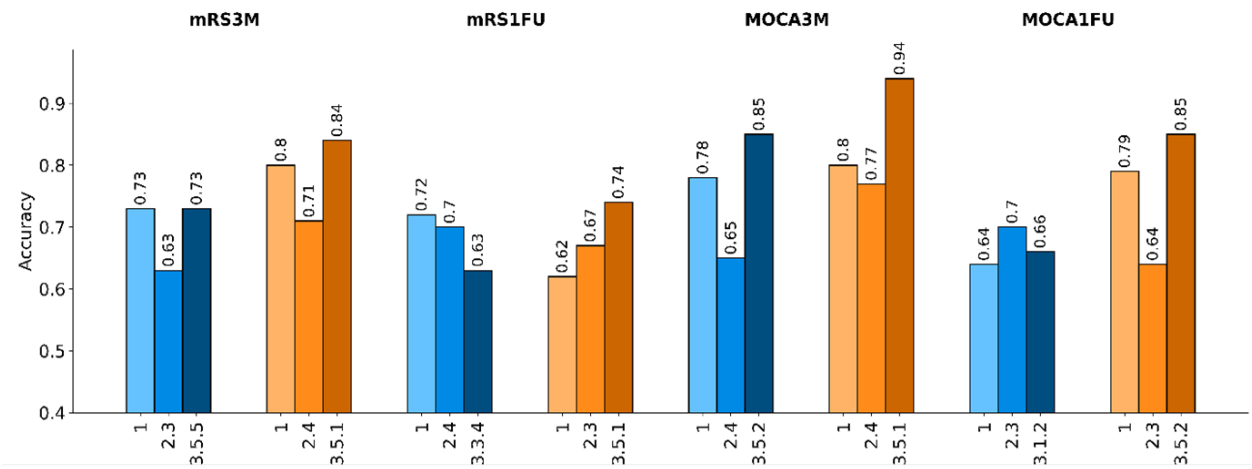

Figure 4.4: Comparison of accuracy metrics among radiomic, deep learning, and hybrid approaches for long-term outcomes

Figure 4.5 showcased the performance improvement of ROC curves for both short-term and long-term clinical outcomes, respectively. Figures 4.5(top) presents ROC curves for the first and last scans of two short-term outcomes, DCI and VPS. From the DCI plots, it is evident that two hybrid approaches had AUC values of 0.91 ± 0.05 and 0.81 ± 0.05. *M1* and *M2* reported AUCs of 0.82 ± 0.05 and 0.79 ± 0.05, which are smaller or closely comparable to the hybrid approaches.

For VPS, the highest AUC value was from *M2*, with 0.92 ± 0.04. Approach 3.5.5 yielded an AUC of 0.86 ± 0.05, which is also significantly higher than approach 1's AUC of 0.77 ± 0.07, indicating that the hybrid approach yields superior or comparable prediction outcomes for all short-term prognoses.

Evaluating the long-term outcomes from figure 4.5(bottom), it was observed that MOCA 3M ranged from 0.58 ± 0.19 to 0.94 ± 0.04, with the highest value attained by approach 3.5.1. *M1* scored 0.86 ± 0.06, while the maximum score for *M2* (2.4) was 0.78 ± 0.012, both of which were surpassed by the hybrid approach, *M3*. Another long-term outcome, mRS1F, also displayed its highest value with approach 3.5.1, with AUC values ranging from 0.54 ± 0.22 to 0.75 ± 0.08. These comparisons highlight the enhanced performance of the model when the hybrid approach is used.

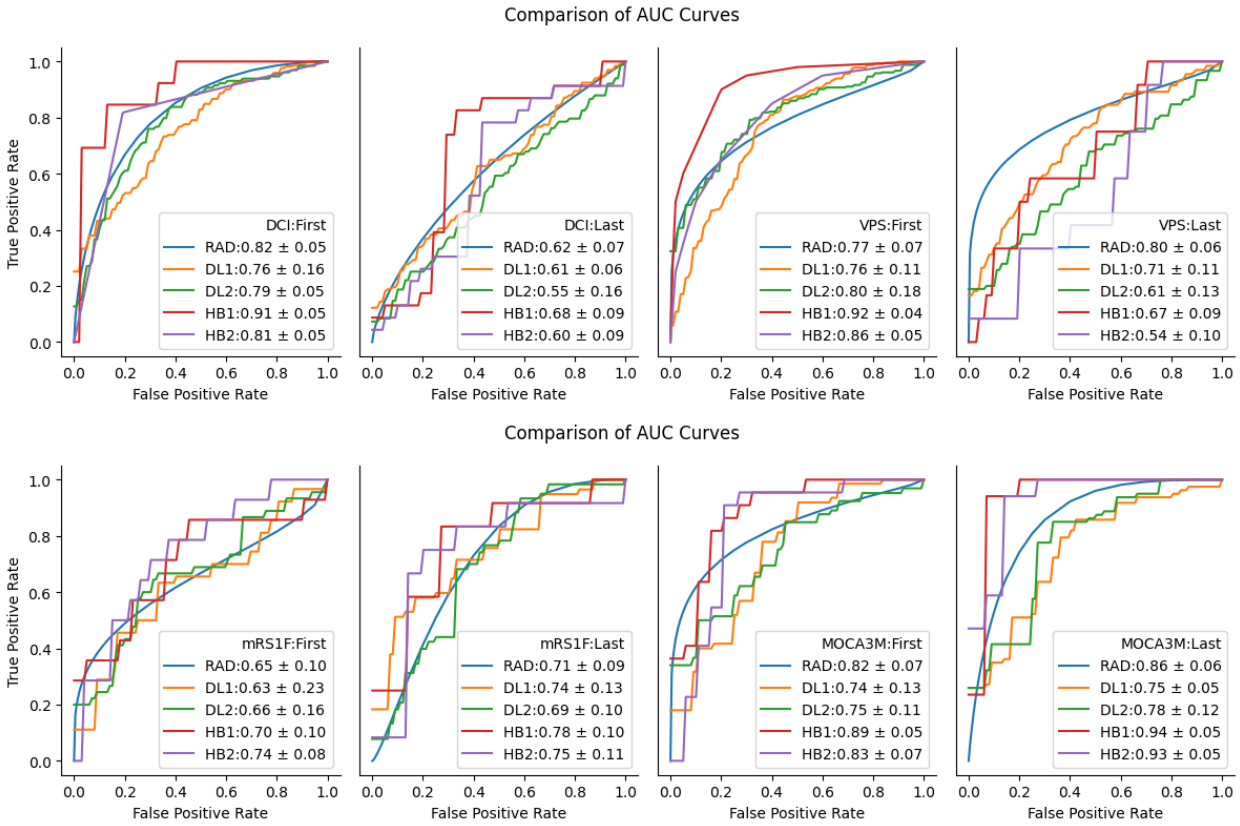

Figure 4.5: A comparison of ROC curves for (top) short-term clinical measure DCI and VPS (bottom) long-term clinical measure mRS1F and MOCA3M

Table 4.4 and 4.5 provide additional comparison results for all three approaches concerning both short-term and long-term clinical measures, respectively. However, in Table 5, the highest sensitivity for HCP (0.89) is observed in approach 2.4, while both approach 3.5.2 and 3.5.3 show a value of 0.8. Yet, upon comparing other metrics of the clinical measure, it becomes evident that 3.5.2 and 3.5.3 offer the best models for HCP. Similarly, for CVSM, *M1* achieves the highest sensitivity of 0.87 as opposed to 0.69 from approach 3.3.2. But comparing other metrics of CVSM, model 3.3.2 offers the best model for CVSM. Thus, by comparing all metrics and overall evaluation, it is clear that the ensemble technique provides the best or at least comparable predictive model for aSAH patient's clinical outcomes.

The model also evaluated the variances in short- and long-term clinical indicators between the admission and discharge scans. In figure 4.5(top) and table 4.4 a comparison of perfor-

Table 4.4: Performance metrics results for short-term clinical measures: radiomic, deep learning, and hybrid approaches

DCI								HCP						
		AUC	ACC	Pre	F1	Sens	Spec		AUC	ACC	Pre	F1	Sens	Sepc
1	First	0.82±0.05	0.78	76	0.81	0.88	0.65	1	0.73±0.09	0.72	0.61	0.67	0.73	0.71
1	Last	0.62±0.07	0.63	0.65	0.72	0.79	0.42	1	0.68±0.09	0.68	0.22	0.26	0.67	0.68
2.1	First	0.71±0.01	0.65	0.70	0.68	0.68	0.61	2.1	0.52±0.14	0.57	0.54	0.43	0.44	0.65
2.1	Last	0.62±0.15	0.60	0.64	0.56	0.51	0.70	2.1	0.46±0.28	0.45	0.22	0.26	0.35	0.51
2.2	First	0.67±0.11	0.60	0.65	0.69	0.82	0.34	2.2	0.37±0.22	0.43	0.38	0.48	0.69	0.27
2.2	Last	0.58±0.13	0.57	0.58	0.67	0.82	0.56	2.2	0.31±0.01	0.38	0.21	0.25	0.38	0.38
2.3	First	0.72±0.01	0.69	0.71	0.74	0.79	0.56	2.3	0.70±0.01	0.61	0.58	0.53	0.58	0.64
2.3	Last	0.67±0.18	0.68	0.70	0.72	0.78	0.53	2.3	0.52±0.15	0.52	0.35	0.42	0.56	0.49
2.4	First	0.69±0.01	0.64	0.63	0.74	0.89	0.33	2.4	0.48±0.25	0.41	0.38	0.54	0.89	0.12
2.4	Last	0.61±0.13	0.58	0.60	0.65	0.72	0.40	2.4	0.47±0.20	0.42	0.24	0.32	0.46	0.41
3.3.1	First	0.81±0.05	0.83	0.81	0.85	0.90	0.73	3.5.2	0.92±0.05	0.90	1	0.88	0.80	1
3.1.1	Last	0.68±0.09	0.63	0.64	0.66	0.69	0.57	3.1.2	0.65±0.08	0.67	0.58	0.56	0.53	0.76
3.3.5	First	0.81±0.05	0.81	0.82	0.83	0.84	0.76	3.5.3	0.85±0.06	0.87	0.85	0.82	0.80	0.91
3.1.4	Last	0.60±0.09	0.61	0.62	0.63	0.65	0.57	3.1.4	0.76±0.07	0.76	0.66	0.71	0.76	0.76
CVSM								VPS						
		AUC	ACC	Pre	F1	Sens	Spec		AUC	ACC	Pre	F1	Sens	Spec
1	First	0.76±0.06	0.78	0.67	0.67	0.87	0.72	1	0.77±0.07	0.77	0.52	0.65	0.85	0.75
1	Last	0.71±0.07	0.63	0.53	0.64	0.83	0.5	1	0.80±0.06	0.73	0.59	0.77	0.77	0.71
2.1	First	0.66±0.13	0.62	0.53	0.51	0.53	0.68	2.1	0.82±0.01	0.77	0.54	0.36	0.36	0.90
2.1	Last	0.56±0.07	0.59	0.42	0.34	0.30	0.76	2.1	0.50±0.22	0.50	0.36	0.44	0.64	0.42
2.2	First	0.66±0.11	0.58	0.48	0.56	0.71	0.50	2.2	0.72±0.05	0.67	0.44	0.50	0.66	0.68
2.2	Last	0.56±0.13	0.42	0.35	0.47	0.76	0.20	2.2	0.63±0.19	0.60	0.33	0.34	0.45	0.66
2.3	First	0.72±0.01	0.66	0.57	0.61	0.70	0.63	2.3	0.89±0.01	0.82	0.65	0.60	0.58	0.89
2.3	Last	0.64±0.11	0.61	0.48	0.47	0.49	0.67	2.3	0.63±0.20	0.68	0.38	0.37	0.42	0.79
2.4	First	0.68±0.10	0.59	0.51	0.61	0.80	0.45	2.4	0.74±0.11	0.70	0.44	0.45	0.54	0.77
2.4	Last	0.56±0.13	0.50	0.43	0.56	0.88	0.25	2.4	0.63±0.01	0.57	0.34	0.45	0.52	0.72
3.3.2	First	0.75±0.05	0.76	0.69	0.69	0.69	0.80	3.5.5	0.92±0.04	0.86	0.68	0.75	0.84	0.87
3.5.1	Last	0.62±0.07	0.62	0.52	0.57	0.63	0.62	3.4.1	0.67±0.09	0.54	0.36	0.48	0.75	0.46
3.4.1	First	0.69±0.06	0.69	0.59	0.64	0.69	0.69	3.3.4	0.86±0.05	0.79	0.56	0.62	0.69	0.82
3.3.4	Last	0.66±0.06	0.64	0.53	0.62	0.73	0.58	3.2.3	0.54±0.10	0.57	0.28	0.30	0.30	0.66

mance metrics for short-term clinical measures can be observed between the scans, which signifies short terms are more accurately predicted using the initial CT scan. For example, HCP exhibits the top AUC score from the initial scan which stands at 0.87, whereas the highest reported from the final scan is 0.73. Similarly, for VPS, an AUC score of 0.92 is noted

Table 4.5: Performance metrics results for long-term clinical measures: radiomic, deep learning, and hybrid approaches

mRS3M								mRS1F						
		AUC	ACC	Pre	F1	Sens	Spec		AUC	ACC	Pre	F1	Sens	Spec
1	First	0.75±0.08	0.73	0.72	0.83	0.96	0.29	1	0.65±0.10	0.71	0.80	0.84	0.91	0.43
1	Last	0.85±0.07	0.80	0.82	0.85	0.88	0.64	1	0.71±0.09	0.62	0.84	0.84	0.80	0.36
2.1	First	0.59±0.22	0.56	0.42	0.49	0.63	0.50	2.1	0.62±0.23	0.62	0.51	0.57	0.68	0.57
2.1	Last	0.82±0.19	0.71	0.66	0.56	0.49	0.84	2.1	0.74±0.13	0.62	0.70	0.48	0.45	0.76
2.2	First	0.58±0.19	0.59	0.41	0.44	0.49	0.64	2.2	0.62±0.18	0.60	0.47	0.51	0.58	0.60
2.2	Last	0.72±0.22	0.63	0.46	0.52	0.64	0.63	2.2	0.54±0.22	0.56	0.51	0.58	0.68	0.58
2.3	First	0.64±0.20	0.63	0.66	0.59	0.73	0.64	2.3	0.63±0.17	0.64	0.54	0.60	0.71	0.56
2.3	Last	0.81±0.19	0.73	0.71	0.61	0.60	0.83	2.3	0.76±0.13	0.67	0.68	0.60	0.76	0.57
2.4	First	0.59±0.18	0.6	0.42	0.48	0.63	0.56	2.4	0.66±0.16	0.70	0.60	0.55	0.58	0.76
2.4	Last	0.81±0.19	0.74	0.66	0.69	0.74	0.74	2.4	0.69±0.10	0.63	0.0.56	0.62	0.70	0.57
3.5.5	First	0.71±0.07	0.73	0.60	0.62	0.64	0.77	3.3.4	0.70±0.10	0.63	0.52	0.60	0.71	0.59
3.5.1	Last	0.85±0.06	0.84	0.73	0.81	0.91	0.80	3.5.1	0.78±0.10	0.74	0.66	0.74	0.83	0.66
3.3.1	First	0.70±0.07	0.70	0.55	0.62	0.71	0.70	3.1.4	0.74±0.08	0.68	0.53	0.55	0.57	0.74
3.1.1	Last	0.84±0.06	0.84	0.76	0.80	0.83	0.85	3.5.2	0.75±0.11	0.70	0.64	0.69	0.75	0.66
MOCA3M								MOCA1F						
		AUC	ACC	Pre	F1	Sens	Spec		AUC	ACC	Pre	F1	Sens	Spec
1	First	0.82±0.07	0.78	0.76	0.81	0.74	0.82	1	0.73±0.08	0.64	0.61	0.67	0.75	0.55
1	Last	0.86±0.06	0.80	0.65	0.71	0.84	0.76	1	0.86±0.07	0.79	0.22	0.26	0.86	0.75
2.1	First	0.65±0.20	0.60	0.66	0.62	0.63	0.57	2.1	0.63±0.10	0.61	0.64	0.68	0.78	0.39
2.1	Last	0.73±0.06	0.72	0.79	0.70	0.67	0.76	2.1	0.51±0.23	0.56	0.67	0.65	0.67	0.33
2.2	First	0.58±0.19	0.61	0.51	0.57	0.68	0.49	2.2	0.51±0.17	0.47	0.50	0.61	0.80	0.08
2.2	Last	0.61±0.15	0.58	0.68	0.063	0.71	0.45	2.2	0.45±0.17	0.63	0.67	0.76	0.89	0.1
2.3	First	0.74±0.13	0.68	0.71	0.70	0.70	0.67	2.3	0.75±0.06	0.70	0.72	0.74	0.78	0.60
2.3	Last	0.75±0.05	0.73	0.78	0.72	0.71	0.75	2.3	0.63±0.10	0.64	0.71	0.74	0.79	0.33
2.4	First	0.75±0.11	0.65	0.67	0.68	0.74	0.56	2.4	0.71±0.14	0.67	0.69	0.73	0.81	0.49
2.4	Last	0.78±0.12	0.77	0.72	0.81	0.94	0.59	2.4	0.47±0.20	0.67	0.68	0.78	0.91	0.13
3.5.2	First	0.89±0.05	0.85	0.80	0.85	0.90	0.73	3.5.1	0.66±0.08	0.66	0.72	0.68	0.65	0.68
3.5.1	Last	0.94±0.05	0.94	0.94	0.94	0.94	0.93	3.5.2	0.84±0.07	0.85	0.88	0.88	0.88	0.80
3.3.1	First	0.83±0.07	0.85	0.80	0.87	0.95	0.73	3.1.2	0.66±0.08	0.66	0.70	0.70	0.70	0.62
3.5.5	Last	0.93±0.05	0.84	0.83	0.85	0.88	0.60	3.1.5	0.79±0.08	0.81	0.83	0.85	0.88	0.70

from the admission scan, surpassing the maximum value of 0.67 from the discharge scan. Table 4.4 illustrates that DCI, HCP, CVSM, and VPS have the best overall performance metrics evaluated from all the admission scans than the discharge scans.

Conversely, the long-term evaluations depicted in figure 4.5(bottom) and Table 4.5 reveal enhanced performance observed in the discharge CT scan compared to the initial one. For example, MOCA3M demonstrates an increase in AUC score from 0.85 to 0.94, while mRS3M reaches a peak of 0.95, compared to 0.75 in the initial scan. Similarly, MOCA3M exhibits higher sensitivity in the final scan (0.94) versus the initial scan (0.87), and mRS3M achieves a sensitivity of 1 in the final scan, contrasting with 0.96 in the initial scan. Considering other performance metrics, table 6 consistently shows that the trend of superior performance from the discharge scan compared to the admission scan remains constant across all four long-term clinical measures.

4.5 Discussion and Conclusion

This study aimed to develop a predictive model based on quantitative imaging biomarkers to classify outcomes in aSAH by assessing and comparing multiple approaches. Eight clinical measures from two CT scans (admission and discharge) were investigated. This included four short-term and four long-term clinical outcomes associated with aSAH, currently assessed by neuroradiologists and clinicians in practice through qualitative assessment and subjective interpretation, potentially resulting in notable inter-rater variability [189]. Although several clinical biomarkers have been proposed for predicting aSAH prognoses, autonomous imaging biomarkers are relatively underexplored [47], [190]. In our initial research, we used a traditional radiomics method to compute features, focusing on the volume of sulci in various brain regions. Then, we expanded our investigation to include DTL methods combined with an attention mechanism for feature extraction. We incorporated complex patterns and abstract features from CT images using pre-trained architectures focusing on blood clot volume in the cisternal space. Thus, in this study, we leveraged the strengths of traditional radiomics and DTL approaches to develop ensemble prediction model-based imaging biomarkers to explore

their applicability in predicting clinical outcomes in aSAH patients. Our study has several novel findings in predicting aSAH outcomes. First, the study validated the effectiveness of quantitative imaging biomarkers in developing an effective prediction model for patients with aSAH. The models' exceptional performance in distinguishing between two groups is demonstrated by their high AUC scores for both short- and long-term outcomes. We examined the short-term results and found that HCP and VPS had AUC values of 0.92±0.05 and 0.86±0.05, respectively. When evaluating the long-term results, mRS3M and MOCA3M had AUC values of 0.85±0.07 and 0.93±0.04, respectively. These indicate how effectively quantitative features perform as imaging biomarkers for prognosticators.

Second, three different methods were evaluated and compared to identify an optimal model for short- and long-term clinical outcomes. Our observation suggested that the hybrid approach offers the best prediction model in aSAH patients. The hybrid method yielded the highest values for most of the performance metrics for DCI. However, the AUC value of the hybrid model for this outcome was still comparable, with a score of 0.91±0.04 compared to 0.82±0.05 from the radiomic approach. For HCP, we noted that the hybrid approach outperforms all evaluation metrics, including an accuracy of 0.90. Moreover, the sensitivity of VPS, which signifies how well the model predicts patients with poor outcomes, is observed significantly better with the same hybrid model (0.84) than the mentioned DL model (0.58). Meanwhile, the AUC using a hybrid approach (0.89±0.01) is slightly lower but comparable with the best-reported score of the attention-based DTL method (0.86±0.05). A similar assessment was observed on long-term outcomes, i.e., MOCA3M performed the best using the hybrid approach (*M3.5.1*) with discharge scan, with an accuracy of 0.94 compared to that of 0.80 and 0.77 from radiomic and DTL models, respectively. Similar trends were observed with 1-year follow-ups where ensemble approaches yielded better performance compared to individual approaches. For instance, MOCA1F reported the highest accuracy of 0.85 for *M3* compared to 0.79 and 0.70 from approaches *M1* and *M2.3*, respectively. All analyses support that adopting the ensemble approach has enhanced the prediction model's performance for all clinical outcomes or yielded similar outcomes compared to the other two approaches.

Moreover, the smaller standard deviation scores of 0.05 for DCI and 0.01 for VPS from the hybrid models indicate narrower confidence intervals, which enhance the interpretability and reliability of the model predictions for users.

Third, the study highlights the critical role of the scan timing in the prognosis prediction. It was observed that a more accurate prognosis for short-term clinical measures can be obtained from the admission CT scan performed on the first hours of symptom onset. On the other hand, the long-term measures show better results from the CT scan performed 10–14 days during discharge. As depicted in Figure 4.5(top) and Table 4.4, all four short-term measures exhibited the highest performance metrics when analyzing the admission scans of aSAH patients. For instance, HCP had the highest accuracy using the admission scan, at 0.90(i.e., when acute HCP is expected to occur) whereas the best value from the discharge scan across all approaches was 0.74 (i.e., acute HCP is expected to have resolved through CSF drainage in the preceding days of hospitalization). Similarly, the best sensitivity value for CVSM with the admission scans was 0.87, whereas the highest from the discharge scan was 0.83. A comparison of precision and F1 scores for all the short-term outcomes in Table 4.4 consistently yields higher results from the admission scans. This implies that utilizing these image characteristics on the admission scan is likely to accurately predict clinical events during hospitalizations, which would help clinicians with insightful information for clinical decision-making to limit the effects of EBI/DCI.

Additionally, as detailed in Figure 4.5(bottom) and Table 4.5, long-term outcomes exhibit better performance evaluation results when the models are processed with the discharge scans. For instance, the ensemble model of the best AUC score of MOCA3M, reaching 0.93 ± 0.04, was generated from the discharge scans, whereas the highest from the admission scan was only 0.84 ± 0.05. Other performance metrics of MOCA3M were also depicted significantly better with the discharge scan than with the admission scan. The only exception observed was in the sensitivity of mRS1F, which had the highest sensitivity coming from the admission scan (0.91), whereas the best from the discharge scan was 0.83. However, all other metrics of this outcome showed higher or comparable performance with the discharge scans.

This observation highlights how patient treatment affects long-term clinical outcomes and provides practitioners with relevant information to help them create the most customized recovery plans for patients who are discharged from the hospital.

Through the evaluation of multiple approaches, the study was able to uncover several favorable results. However, we also acknowledge several limitations of our study. First, there was a possibility that the small dataset size might limit the variety of representations it included. Future research should prioritize growing the dataset size because it will enable the addition of more characteristics and enhance the precision of performance model predictions, leading to a more reliable CAD system. Second, we have only explored a small subset of radiomic features (sulci) in our analysis; more exploratory analysis to extract other types of radiographic information still needs to be investigated. Third, we have investigated several other state-of-the-art DTL methods in our analysis but were not able to achieve a significant performance boost given the limited dataset size; with large and diverse datasets, we can continue to explore other DTL variants. These future models have a high potential to perform even better if additional feature engineering and DTL techniques are implemented.

In summary, despite the limitations with dataset size, the study effectively demonstrated the comparison among radiomics, deep learning, and ensemble approaches for developing prognostic prediction models for aSAH patients. After a comprehensive analysis and discussion, it was determined that the ensemble approach, which combined the advantages of deep learning models and radiomics, performed the best overall out of all the models. The greater AUC score, accuracy, sensitivity, and F1 score showed how well the model identified patients who could have excellent outcomes. This makes it possible for clinicians with an objective toolkit to counsel patients and family members during hospitalization. Our study also highlights the differences in using CT scans at different time points during hospitalization may alter the predictability of various outcomes. Hence, researchers and clinicians need to be cognizant of such disparities while predicting outcomes to counsel patients or family members. Building on the groundwork provided through this work, future research endeavors can concentrate on improving and verifying these new quantitative image indicators and/or

including additional clinical or laboratory metrics using prospective clinical studies.

CHAPTER 5

Develop Composite Radiological Tool to Predict Clinical Outcomes after Ischemic Stroke Treatment

5.1 Introduction

AIS is caused by blood clots obstructing brain arteries that lead to reduced blood flow and oxygen in certain areas, resulting in severe neurological damage or death. In 2022, over 7.6 million new AIS cases were reported globally, accounting for more than 62% of all strokes. Approximately 3.3 million of these cases resulted in death, with stroke-related fatalities and disabilities contributing to over 63 million disability-adjusted life-years (DALYs) annually[136, 45].

The functional outcome of a patient with AIS is predominantly shaped by their initial clinical presentation and the extent of brain tissue damage [161, 191]. A robust predictive model is essential for assessing treatment effects, follow-up care, and making timely adjustments to the patient's social environment, such as arranging early long-term care[192]. The low blood volume in the brain within the initial few hours following a stroke causes

early signals of the proliferation of cells on radiology scans, which may indicate the possibility of permanent damage soon[193, 194]. The European Cooperative Acute Stroke Study (ECASS) conducted two key trials that highlighted an important link between early stroke changes and outcomes in acute ischemic stroke. The trials showed that prognosis depends on whether more or less than one-third of the middle cerebral artery (MCA) area is involved [162]. However, accurately identifying and quantifying these areas had been challenging, even for experienced stroke physicians and radiologists.

Later, ASPECTS was developed to tackle the limitation as a reliable imaging tool for predicting functional outcomes. This scoring system was initially used to qualitatively assess symptomatic hemorrhage following intravenous thrombolysis and has become widely utilized in clinical practice to evaluate early ischemic changes in the MCA territory [52]. ASPECTS was initially developed for non-contrast CT scans as a 10-point topographic scoring system, later adapted for use with DWI-MRI scans to improve sensitivity [195]. This method divides the MCA territory into 10 regions, assigning 1 point for each region with normal-appearing tissue and deducting 1 point for each region showing ischemic damage. Patients with ASPECTS scores < 7 were identified as higher-risk patients. To address the prevalence and unique challenges of posterior circulation infarctions, the ASPECTS scoring system was expanded with an additional 10 points and named PC-ASPECTS [196]. These strokes from the posterior circulation infarcted area represent a significant subset of acute ischemic cases, affecting critical motor and sensory pathways that are essential for functional recovery. Although the initial ASPECTS studies focused on identifying early ischemic changes, subsequent research demonstrated that post-ASPECTS and post-PC-ASPECTS scores, measured after thrombectomy, can predict functional outcomes [197, 198].

Although ASPECTS is favored in clinical practice for its simplicity, conflicting findings have surfaced regarding its inter- and intra-observer variability. The agreement is generally strong for patients with higher ASPECTS scores (> 7), but the reliability decreases significantly for patients exhibiting extensive early ischemic changes [55]. Outcome variations based on the affected area have also been observed in ASPECTS regions from baseline CT

scans [199, 200]. The traditional ASPECTS assigns equal weight to all brain regions, which reduces its predictive accuracy, as research suggests that not all ASPECTS regions hold the same prognostic value for predicting modified Rankin Scale (mRS) scores, a measure of cognitive dysfunction, at three months. As a result, some ASPECTS regions contribute minimally to mRS score predictions [166]. Subsequent studies have indicated that ASPECTS blood volume is an additional effective imaging biomarker for outcome prediction beyond the traditional ASPECTS system [174, 167]. Recent studies have also used ischemic volume as a primary feature for outcome prediction [172, 201]. To account for variations in brain size—especially given that stroke size can significantly impact brain volume and that brain size varies widely among children of different ages—it is more accurate to express infarct volume as a percentage of total brain volume [54, 173].

Figure 5.1 illustrates the central aim of my third study by highlighting the limitations in current protocols for assessing prognosis in AIS patients. The visual representation outlines a typical patient's treatment protocol from the initial symptom onset; screening methods; treatment procedures; and prognostic assessment. ASPECTS/PC-ASPECTS score and the National Institutes of Health Stroke Scale (NIHSS), were conducted pre- and post-treatment, which guides doctors in treatment planning and assessing potential outcomes. mRS score is also collected at discharge and the 3-month follow-up. Even though multiple radiological examinations were acquired, only post-treatment scans were used in our analysis. As mentioned earlier, due to the qualitative nature of the ASPECTS score assessment, it is still unclear whether each of these regions should be weighted equally, or whether quantitative assessment of the CIV volumes with respect to the ASPECTS and PC-ASPECTS regions would be a composite tool for prognostic assessment.

Moreover, while ASPECTS has proven useful, recent advances in AI and ML are set to transform stroke diagnosis and prognosis. A review of AI and ML in medical imaging shows that these technologies can improve the precision and consistency of stroke assessments by automating the analysis of complex imaging data [168]. Using these advanced methods can address the limitations of subjective assessments, leading to more accurate and reliable

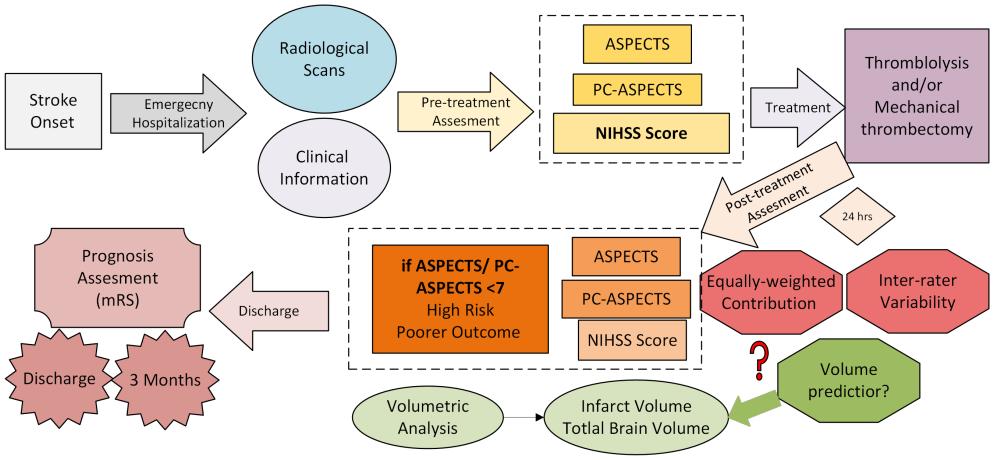

Figure 5.1: Traditional AIS patient diagnostic and evaluation process: limitations and potential improvements

predictions of patient outcomes.

Thus, based on the background, my third study tests the feasibility of developing an optimal prognostic model to predict functional outcomes in AIS patients using a comparative volumetric analysis using multiple machine learning algorithms. The study evaluates infarcted tissues across 20 regions defined by ASPECTS and PC-ASPECTS. Specifically, it examines the ratio of cerebral ischemic volume (CIV) in ASPECTS and PC-ASPECTS regions, derived from post-intervention MRI and CT scans (performed 24 h of AIS onset), relative to the total brain volume (TBV), treating these ratios as potential imaging biomarkers. These biomarkers are analyzed in combination with pertinent clinical characteristics of the patients. Notably, this study represents the first predictive investigation of functional outcomes in AIS patients that integrates volumetric information from both anterior and posterior circulations in relation to overall brain volume.

5.2 Framework of Proposed Prognosis Model:

The proposed prognosis model centers on volumetric analysis to predict functional outcome mRS score at discharge and at 3 months. Figure 5.2 provides an overview of the study's

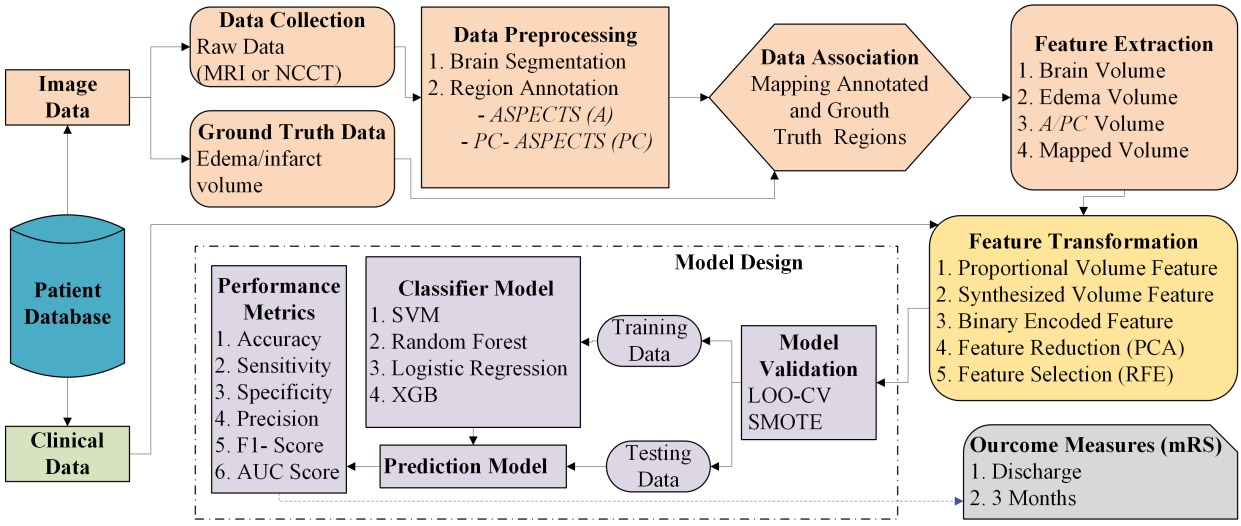

Figure 5.2: Framework of the proposed model for predicting functional outcomes of AIS patients

methodology, starting from data collection through performance evaluation of the model. Radiographic data (MRI or CT) underwent image processing, with ASPECTS(AP) and PC-ASPECTS(PA) regions marked and mapped against ground truth data, consisting of CIV. Image features were extracted and tailored to study requirements. Clinical data were transformed into a format usable by the model. The transformed feature data, alongside labeled data, underwentnt additional feature engineering before being fed into the model. Validation was conducted using LOOCV to maximize data efficacy. To address the class imbalance, SMOTE was applied. The model employed multiple classifier algorithms, which were compared and analyzed to propose the best predictive model and evaluated through multiple performance metrics.

Table 5.1: Dataset size (total cases/by class)

mRS Assessment Time	Total number of cases	Class 0	Class 1
Discharge	179	49	130
3 Months	167	77	90

5.3 Methodology

5.3.1 Data Collection

This retrospective study was approved through an agreement between the Institutional Review Board of the University of Oklahoma Health Sciences Center and UT Southwestern. Data collection, identification, and de-identification were exclusively conducted at UT Southwestern by one of their neurologists, while the analytical process, including the development of a predictive model using the de-identified data, was carried out at the University of Oklahoma's Norman Campus. The study included individuals admitted to Clements University Hospital (CUH) and Texas Health Presbyterian Hospital (THD) with a confirmed diagnosis of AIS who received thrombolysis treatment (using either ALT or TNK) between January 2019 and December 2022. Eligible patients were identified through the prospective stroke database maintained at CUH and THD. All relevant data were securely stored in a HIPAA-compliant online REDCap database. The dataset consisted of 179 patients and included either non-contrast CT images or axial diffusion-weighted images (DWI) with an apparent diffusion coefficient (ADC) map, along with an axial T2/FLAIR MRI sequence. The dataset also included stroke ground truth markings created with itk-SNAP, clinical indicators, and patient outcomes. The outcomes were measured with two key indices: the Modified Rankin Score (mRS), which ranges from 0 (no disability) to 6 (severe disability). Outcomes were assessed through two evaluations: during discharge and at 3-month (3M) follow-ups. Table 5.1 shows the total number of cases for both mRS at discharge and 3 Months and also highlights the class distribution between the positive/poor class (class 1) and negative/good class (class0)

5.3.2 Data Preprocessing

5.3.2.1 Image Data:

A CAD system was developed, illustrated in Figure 5.4, to facilitate region annotation in both CT and MRI scans. For the annotation of ASPECTS/PC-ASPECTS regions, T2/FLAIR, DWI, or CT scans were utilized.DWI sequence has the highest sensitivity, while the T2/FLAIR sequences have a superior anatomical resolution. Expert doctors carefully examined each case using ITK-SNAP to perform CIV ground truth marking in DWI, apparent diffusion coefficient (ADC) maps, or CT scans. Before annotating the ASPECTS and PC-ASPECTS regions, raw radiological data were segmented based on a two-step mapping-based algorithm. Otsu thresholding was used to identify the brain regions, and an estimated center position was set for each image slice, adjusting as each slice was processed to enhance accuracy. The slices were then processed in both forward and backward directions to ensure consistency, followed by morphological operations to smooth edges and fill gaps, resulting in cleaner and more accurate segmentation across all slices. Total brain volume was extracted from the CAD system after the final segmentation.

Then the CAD system was employed to annotate ten anterior regions in two slices as follows: M1, M2, M3, M4, M5, M6, I (insular ribbon), IC (internal capsule), L (lentiform nucleus), and C (caudate) [162]. Ten posterior regions were also annotated in three slices, including the left and right sections of the midbrain, thalamus, occipital lobes, pons, and cerebellar hemisphere [196]. The CAD system utilized a guided image as shown in figure 5.3 to mark the ASPECTS/PC-ASPECTS region. CT was used for the above feature extraction when the above MRI scans were unavailable. All annotated scans were then mapped to scans containing ground truth edema information. All annotated scans were then mapped to infarct volume data. The imaging scan slices ranged from 30 to 36 with average brain and infarct volumes of 1312.63 ± 136.63 mm^3 and 60.01 ± 77.38 mm^3 respectively.

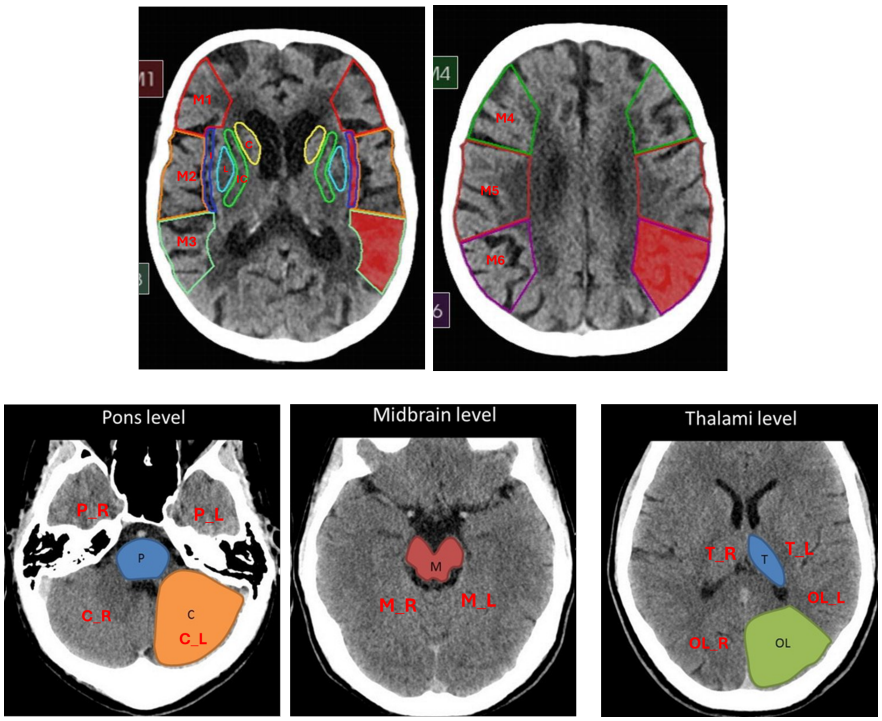

Figure 5.3: Representative radiological slices used for ASPECTS and PC-ASPECTS annotation: (top) two slices illustrating ASPECTS regions; (bottom) three slices for PC-ASPECTS regions

5.3.2.2 Clinical Data

A comprehensive collection of baseline epidemiological and clinical data was extracted as clinical features for the study. Demographic variables, including age and sex, were documented alongside clinical metrics such as body mass index (BMI), diabetes status (HbA1C and glycemic gap), and initial NIHSS scores upon admission. Detailed time data related to thrombolysis and mechanical thrombectomy procedures were also recorded. The study included diverse etiologies of AIS patients and prevalent comorbidities like hypertension and coronary heart disease (hyperlipidemia and atrial fibrillation). Post-interventional clinical parameters specifically changes in NIHSS scores within 24 hours following treatment and the modified Rankin Scale (mRS) scores at discharge and three months, were evaluated to comprehensively assess patient outcomes and treatment efficacy.

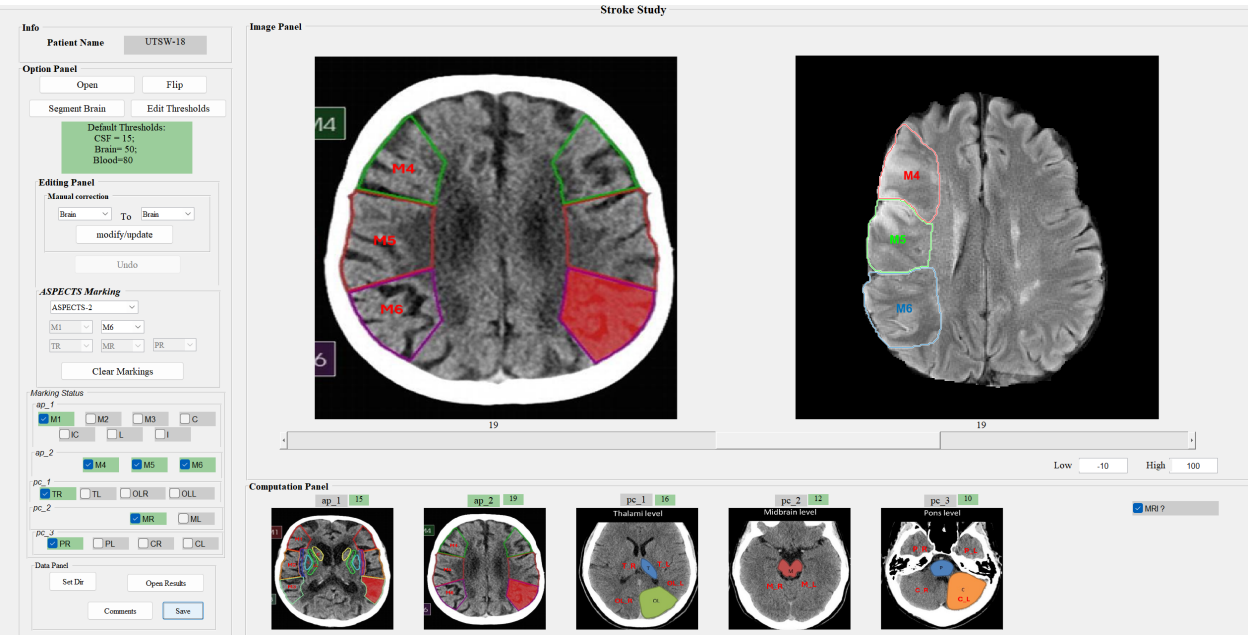

Figure 5.4: CAD system to annotate ASPECTS and PC-ASPECTS region

5.3.3 Feature Engineering

Extracted clinical features were standardized to ensure compatibility with the modeling process. Sex was encoded as 0 for females (F) and 1 for males (M). NIHSS scores were recorded at admission and 24 hours post-intervention, categorized as follows: 0: None-Minimal (scores 0-4), 1: Mild (scores 5-9), 2: Moderate (scores 10-14), 3: Moderately Severe (scores 15-19), and 4: Severe (scores $>$ 20. Detailed timing data for thrombolysis and mechanical thrombectomy procedures were recorded and classified according to [202] as followed:55-129 minutes, 130-204 minutes, and 205-280 minutes. Functional outcomes were assessed using the modified Rankin Scale (mRS) at discharge and three months post-discharge, with scores classified as 'favorable'/'class1' (≤ 2) or 'unfavorable'/'class 2'(> 2).

The proposed CAD system computes quantitative radiomic features from imaging data related to TBV, CIV, and Aspects/PC-ASPECTS raw volumes (20 values). It also computes mapped volumes (20 values), which represent the overlap/intersection of volumes between CIV and Aspects/PC-ASPECTS regions. The above radiomic features are part of the first feature set, FS_1. Additional transformations were done utilizing the mapped volume to

calculate an additional volumetric set of features as part of FS_2. FS_2 included individual proportional volumes (VP_i) calculated using equation 5.1; and sectional volumes VS_{AP}, VS_{PA} and VS_T w.r.to ASPECTS, PC-ASPECTS, and combined regions, respectively, calculated using equation 5.2. Lastly, the presence of CIV in ASPECTS/PC-ASPECTS regions is also created as a binary flag, as depicted in 5.3.

$$VP_i = \frac{Mapped_{Vol}(i)}{CIV} \tag{5.1}$$

$$VS_{Type} = \frac{\sum_{i=x}^{y} Mapped_{Vol}(i)}{CIV} \begin{cases} x = 1, y = 10 & \text{if Type} = \text{AP} \\ x = 11, y = 20 & \text{if Type} = \text{PA} \\ x = 1, y = 20 & \text{if Type} = \text{T} \end{cases} \tag{5.2}$$

$$\sum \text{if } Mapped_{Vol}(i) = 0 \tag{5.3}$$

Table 5.2 summarizes the extracted clinical and radiomic features of FS_1 and FS_2 obtained from the clinical and radiological data.

Table 5.2: Details about the proposed feature set and transformations

Feature Set	Features and/or transformations
FS_1	TBV, CIV, AP/PA raw volume, mapped volume, transformed clinical features
FS_2	VP_i, VS_{AP}, VS_{PA}, VS_T, sum of mapped volume, transformed clinical features

5.3.4 Proposed Prediction Model Design and Evaluation

The two feature sets were used to develop several classification models to assess good and poor prognoses. Each model generates classification probability scores ranging from 0 to 1, with higher scores indicating a greater likelihood of a poor prognosis (positive class: mRS = 3-6). For multi-feature fusion classification models, four well-known architectures

were selected: SVM, Random Forest (RF), Logistic Regression (LR), and Extreme Gradient Boosting (XGB).

In building optimal models, several key steps were taken. First, feature-wise standardization was applied to normalize each attribute. Next, four feature engineering techniques, along with using all raw features, were implemented for both feature sets, as summarized in Table 5.3. The techniques included (i) using all raw features, (ii) reducing feature dimensionality with PCA to retain 95% variance (with 30 features selected using the elbow method), and three feature selection methods: (iii) recursive feature elimination (RFE), (iv) Lasso [203], and (v) Elastic Net (EN) [204].

To address the limited dataset size, a LOCO-CV method was employed to train and evaluate each model, maximizing training samples and reducing case partition bias [205]. Finally, to counter class imbalance, SMOTE was applied during each LOCO training step, generating synthetic samples in the minority class based on the true sample distribution to strengthen training and validation [72].

Table 5.3: Summary of the feature combinations for each ML model

ML Model	**All**	**PCA**	**RFE**	**Lasso**	EN
SVM	SVM	SVM_PCA	SVM_RFE	SVM_Lasso	SVM_EN
LR	LR	LR_PCA	LR_RFE	LR_Lasso	LR_EN
XGB	XGB	XGB_PCA	XGB_RFE	XGB_Lasso	XGB_EN
RF	RF	RF_PCA	RF_RFE	RF_Lasso	RF_EN

To evaluate the performance of each classification model, we used two approaches. First, ROC is constructed using classification scores, and the AUC is computed to assess, evaluate, and compare the performance to classify the mRS discharge. Second, an operating threshold ($T = 0.5$) was used on classification scores to divide all testing cases into two mRS classes (score ≤ 0.5: good prognosis; score > 0.5: poor prognosis). Furthermore, accuracy, sensitivity, specificity, F1-score, and precision, were selected as performance metrics to evaluate the model. Additionally, to attain the statistical significance of the models, the study employed

ROCKIT software which applies a binormal ROC model to get the t-test value assuming the scores for positive and negative cases follow normal distributions.

5.4 Results

Figure 5.5 compares the performance metrics of several classifier models for predicting mRS at discharge using FS_1 with five combinations defined above in table 5.3. With FS_1, accuracy ranged from 0.63 to 0.80, with SVM-RFE and LR-Lasso achieving the highest scores. Notably, PCA did not positively impact performance, whereas all three feature selection techniques (RFE, LASSO, and EN) contributed to improvements. Among the RFE models, RF-RFE achieved the highest precision (0.96) and specificity (0.93), while XGB-RFE demonstrated a strong balance between sensitivity (0.84) and specificity (0.87) alongside high accuracy (0.80) and precision (0.87)—all crucial indicators for a reliable prognostic model. Examining the models utilizing the Lasso technique, SVM-Lasso delivered optimal values in precision (0.96), sensitivity (0.78), and specificity (0.71). The EN models collectively performed exceptionally well, with sensitivity, precision, and accuracy ranging from 0.75 to 0.88, 0.82 to 0.92, and 0.76 to 0.78, respectively.

Figure 5.6 presents a comparative analysis of models for predicting mRS at discharge using FS_2 across various feature combinations. Similar to FS_1, FS_2 showed that adding PCA for dimensionality reduction had a neutral or negative effect, while RFE and EN generally contributed more positively. Accuracy ranged from 0.54 to 0.93, with the lowest score reported by LR-PCA and the highest by the SVM-RFE model. The LR-Lasso model achieved the highest sensitivity (0.94), followed closely by SVM-EN and SVM-RFE, both with scores of 0.91. Overall, LR-RFE and SVM-RFE demonstrated the strongest performance across metrics, with accuracy (0.89, 0.93), sensitivity (0.88, 0.91), precision (0.97, 0.97), and specificity (0.93). Additionally, SVM-EN emerged as a high-performing model with metrics of 0.85, 0.91, 0.93, and 0.86 for accuracy, sensitivity, precision, and specificity, respectively. Overall, FS_1 tends to be a better pick compared to FS_2. Also, LR and SVM with RFE and EN performed significantly better than other approaches, while RF and XGB variations for

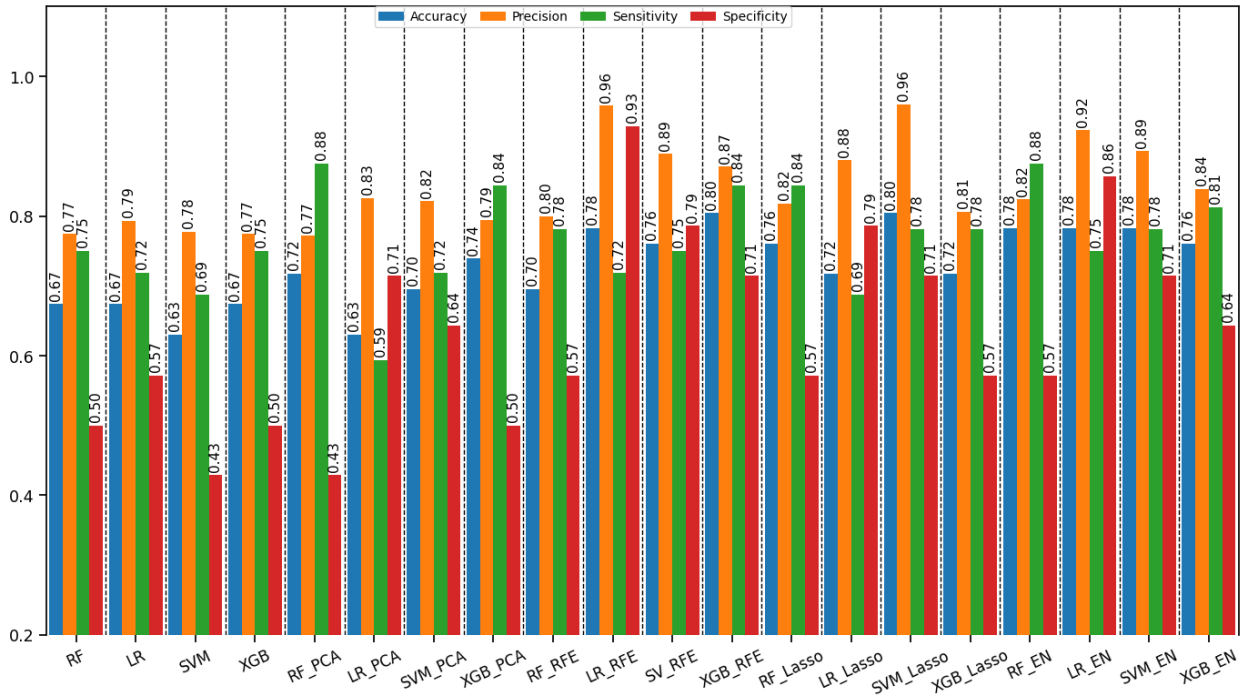

Figure 5.5: Comparison analysis for mRS (discharge) among four ML models with different features using FS_1

both techniques performed the worst.

Figures 5.7 displays the ROC curves for mRS (discharge) analysis, with AUC scores for FS_1 and FS_2 presented at the top and bottom, respectively. The highest AUC of 0.94 for FS_2 was achieved with the SVM and LR models using EN, followed closely by LR-RFE and SVM-RFE (0.93). For FS_2, LR-RFE and SVM-RFE also performed best, with AUC scores of 0.91 and 0.86, respectively. Overall, SVM-All and LR-PCA with FS_1 and FS_2 had the lowest AUC (0.63). In summary, RFE and EN approaches provided the best outcomes, while PCA showed the weakest performance. Among all feature selection techniques, EN consistently delivered the highest AUC scores, while LR models using SVM and RFE techniques also yielded significantly better results. In overall comparison, FS_2 emerged as a better choice than FS_1 for prognosis assessment.

Next, the comparative analysis of models predicting mRS at the 3-month follow-up is presented in Figures 5.8 and 5.9, using FS_1 and FS_2 across five feature combinations. With

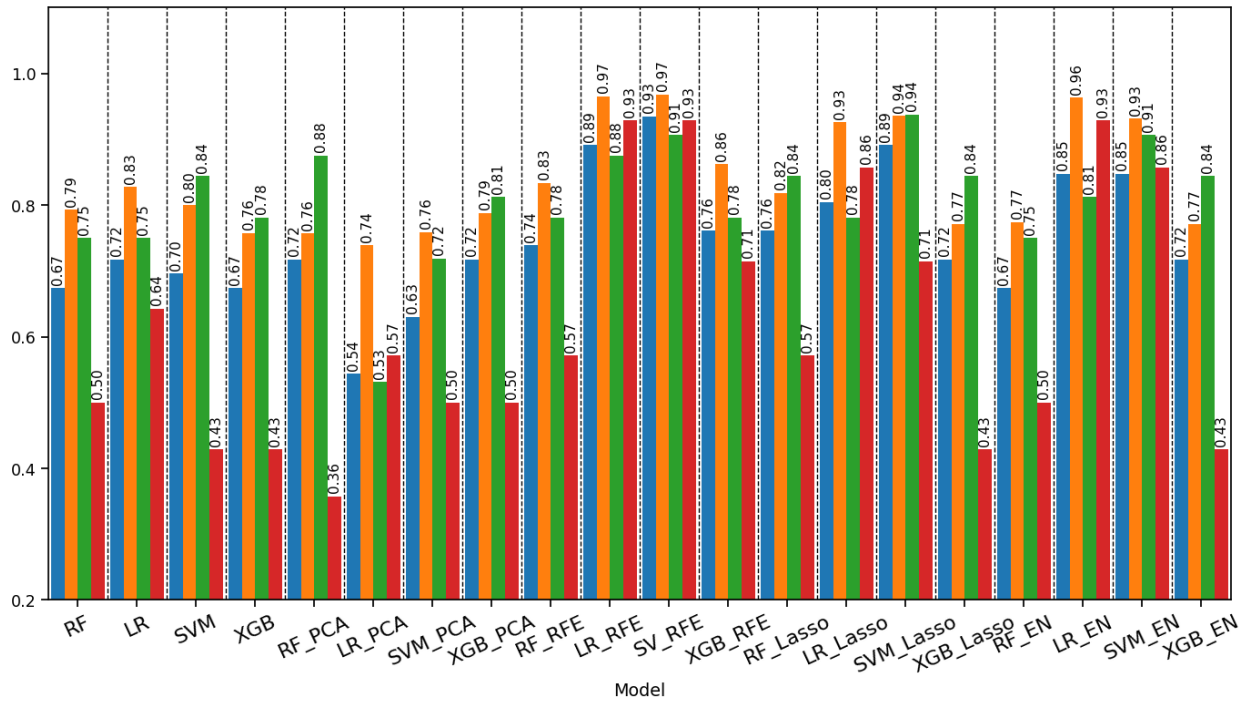

Figure 5.6: Comparison analysis for mRS (discharge) among four ML models with different features using FS_2

FS_1, the highest sensitivity (0.90) was achieved by SVM-RFE, while the lowest score (0.60) was reported by RF-PCA. For FS_2, the EN model outperformed all others, with SVM-EN achieving the highest scores across all metrics: sensitivity (0.88), accuracy (0.84), specificity (0.83), and precision (0.82). Consistent with previous findings, PCA performed poorly with both feature sets. Lasso feature selection technique performed better with FS_1 than FS_2, with the highest accuracy (0.85) seen from RF-Lasso with the first set. Overall, both EN and RFE approaches yielded better performance for FS_1 and FS_2, with EN showing a slight advantage.

Lastly, the ROC analysis with AUC score comparisons for mRS at 3 months in figure 5.10 showed that models using FS_2 performed slightly better than those with FS_1. The highest AUC with FS_2 was achieved by the SVM-EN model (0.93 ± 0.02), while the top performance with FS_1 was (0.91 ± 0.02). RFE also delivered strong results, with both the XGB and SVM models reaching AUCs of (0.92 ± 0.02) and (0.91 ± 0.02), respectively. Overall, EN and

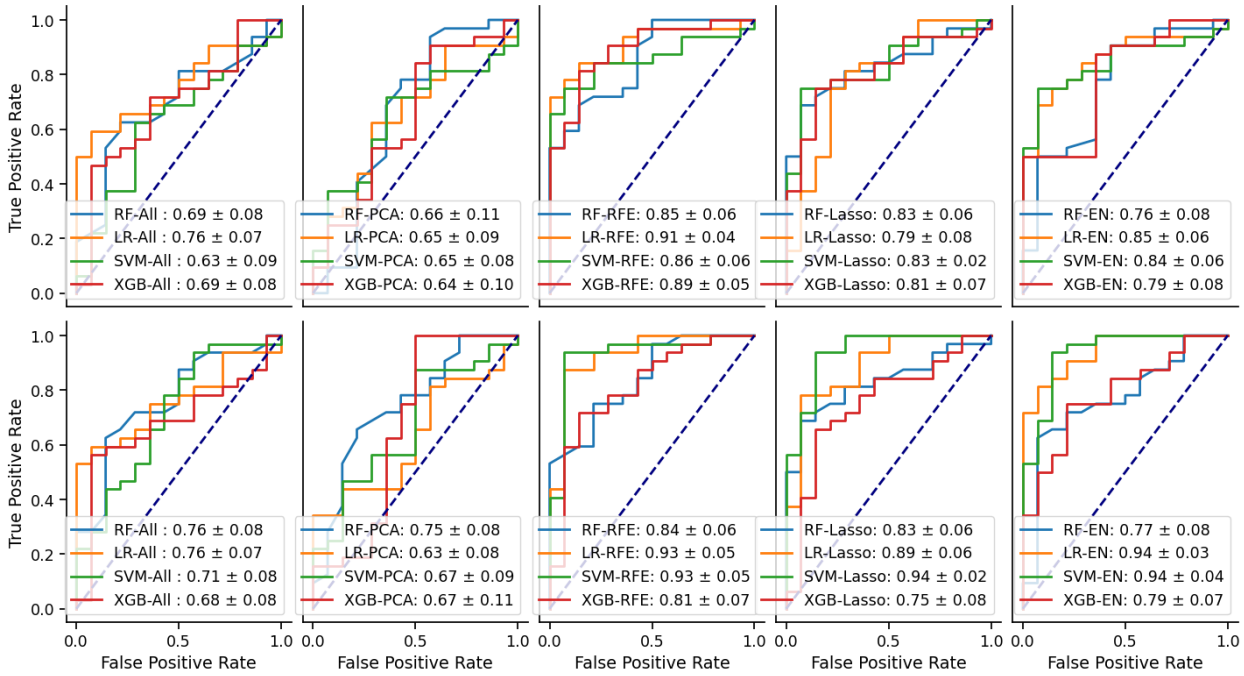

Figure 5.7: ROC analysis comparing ML Models for mRS (discharge) with (top): FS_1 and (bottom): FS_2

RFE with FS_2 were the most effective in predicting 3-month mRS outcomes.

Additionally, the statistical significance of the models was evaluated, revealing that most models achieved statistical significance with p-value less than 0.05, even when their AUC scores were close or similar. Table 5.4 presents the t-test values and p-values for the models with the highest AUC scores predicting mRS at 3 months.

Table 5.4: Statistical significance between selected FS_1 and FS_2 models for mRS 3M

FS_1 Models	FS_2 Models	t-test score	p-value	Statistical Significance
SVM-EN	SVM-EN	-1.63	0.05	significant
LR-RFE	SVM-EN	-2.09	0.02	significant
LR-RFE	SVM-RFE	-0.58	0.27	not significant

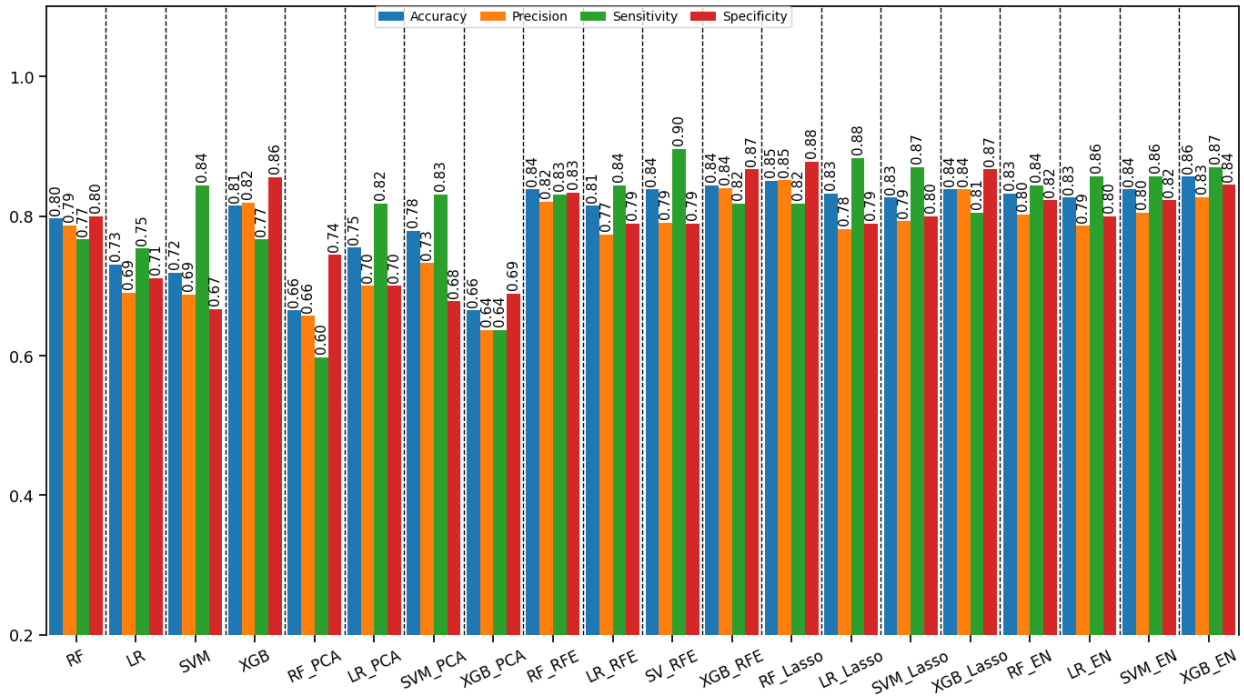

Figure 5.8: Comparison analysis for mRS (at 3 Months) among four ML models with different features using FS_1

5.5 Discussion and Conclusion

The traditional approach to prognostic prediction for ischemic stroke is limited by the unequal weight of AP/PA regions in qualitative assessment along with observer variability [55, 54]. To address these limitations, this study explored the use of a quantitative approach, integrating volumetric information from clinically relevant regions with clinical attributes, to develop a composite tool serving as an imaging biomarker for ischemic stroke patients. This method considered the proportion of affected ASPECTS and PC-ASPECTS regions with respect to the overall brain size, acknowledging their crucial role in accounting for varying stroke and brain sizes.

The annotated ASPECTS and PC-ASPECTS regions, along with the edema volume marked by the physician, were mapped, and volumetric information for the brain, edema, ASPECTS, PC-ASPECTS, and associated mapped regions was extracted as part of FS_1.

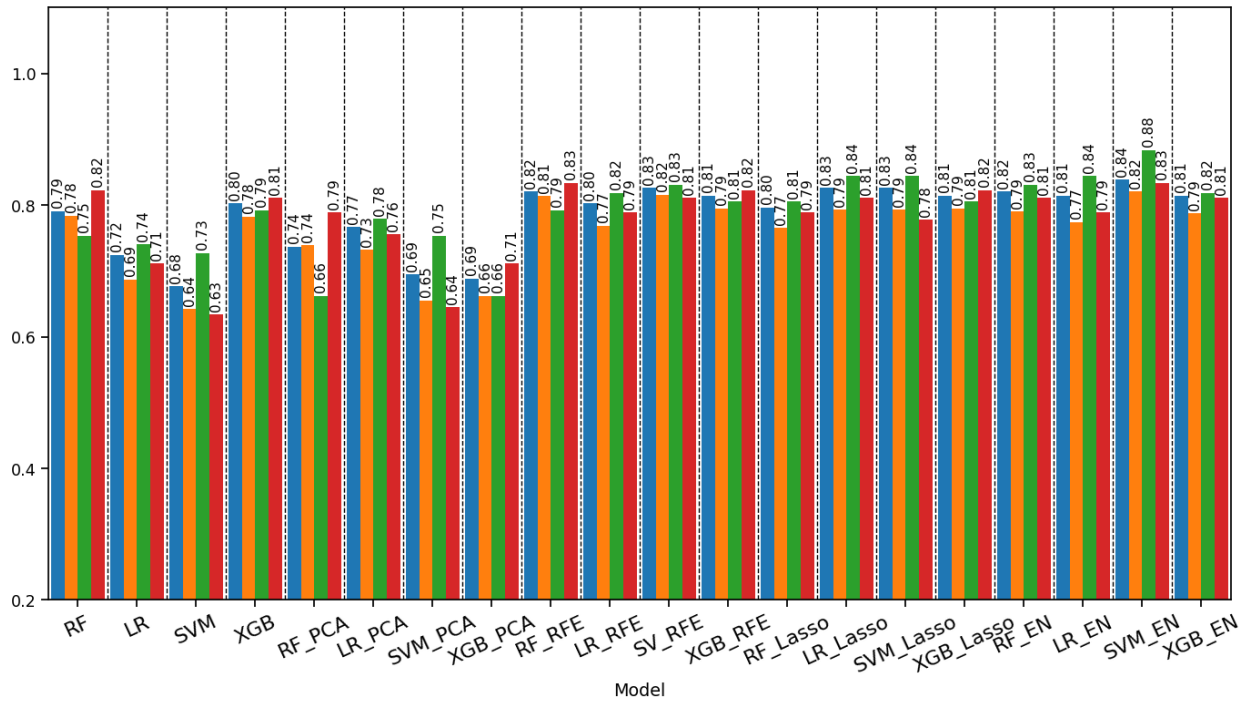

Figure 5.9: Comparison analysis for mRS (at 3 Months) among four ML models with different features using FS_2

Proportional volumes for each ASPECTS and PC-ASPECTS region were calculated relative to total brain volume, while sectional volumes for ASPECTS, PC-ASPECTS, and their combined regions were measured and included in FS_2. Clinical features were also extracted, standardized, and incorporated into both feature sets. Each set was then analyzed separately using four machine learning classifiers—SVM, LR, XGB, and RF. To identify the optimal model, feature engineering techniques including PCA, RFE, LASSO, and Elastic Net were applied, and comparative results were reported. SMOTE and Leave-One-Out Cross-Validation (LOO-CV) were used to enhance performance and address class imbalance. This study makes several notable contributions to ischemic stroke prognosis, underscoring the significance of the research.

First, the study demonstrated the effectiveness of quantitative features in the ASPECTS and PC-ASPECTS regions for predicting functional outcomes. High AUC scores highlight the model's strong ability to distinguish between favorable and unfavorable prognoses. The

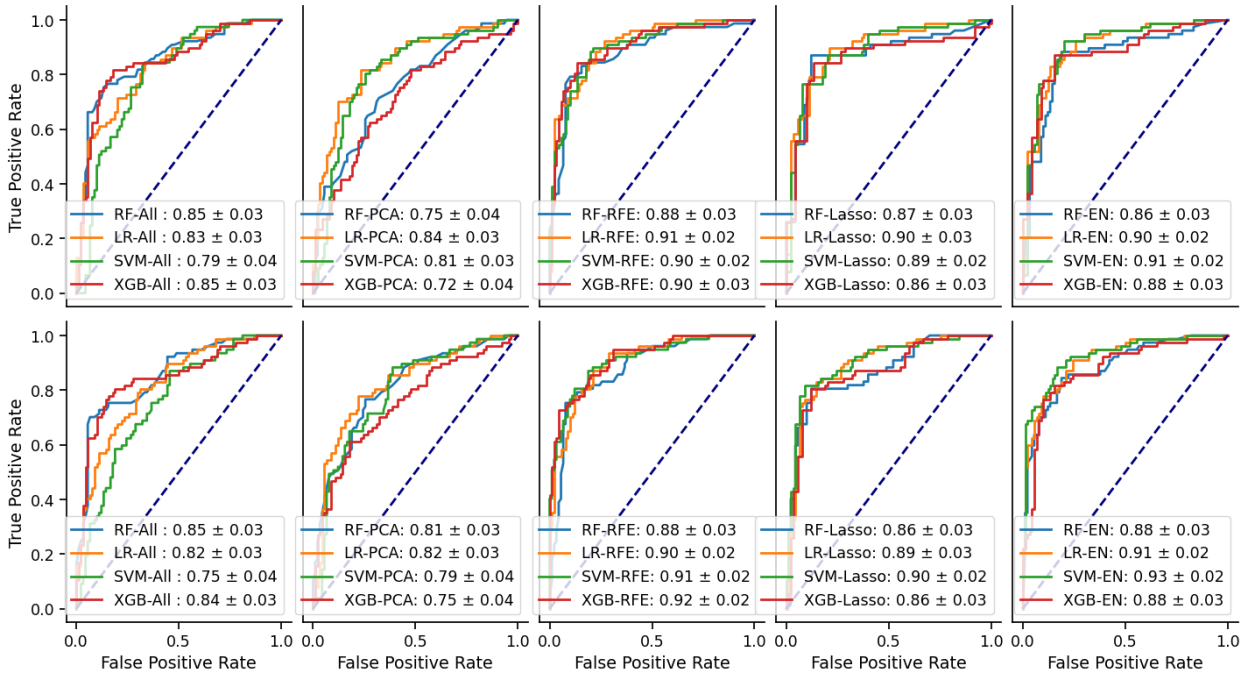

Figure 5.10: ROC analysis comparing ML Models for mRS (at 3 Month) with (top): FS_1 and (bottom): FS_2

models showed exceptional performance in this area, achieving a peak AUC score of 0.93 ± 0.02, underscoring the high potential of volumetric features as a valuable prognostic tool.

Second, an extensive extraction and analysis of multiple volumetric features were conducted to assess their significance. Figures 5.5, 5.6 and 5.7 illustrate the comparison of performance metrics and AUC scores between models utilizing raw features while figures 5.8, 5.9 , and 5.10 present the analysis with transformed features. It was observed that the transformed feature set ($FS_{2)}$ which includes the proportional volume of all the mapped regions along with the sectional volumes and clinical features, showed slightly better performance compared to the raw extracted feature and clinical feature set ($FS_{1)}$. Overall, the analysis highlighted the enhanced predictive capability of the quantitative approach.

Lastly, the study employed and compared multiple classification algorithms and applied various feature engineering techniques to identify the optimal prognostic model, further validating the approach's efficacy. Among all models, PCA showed comparatively lower

performance. The three feature selection techniques—RFE, LASSO, and Elastic Net—all performed exceptionally well, with RFE and Elastic Net demonstrating slightly better and more balanced results. Although the comparisons among these techniques were close, the higher accuracy (0.89) and precision (0.97) underscored the effectiveness of using volumetric features as biomarkers for ischemic stroke prognosis. Moreover, the statistical significance of the models was tested, which was confirmed with p-value less than 0.05 for most cases. As the same dataset was employed for analyses, paired-data comparisons were performed, offering greater sensitivity by accounting for within-sample differences. Even though AUC scores were close, e.g.: (0.93 ± 0.02 VS 0.92 ± 0.02) for the SVM-EN model for FS_2 and FS_1respectively, consistent trends where one model slightly outperformed the other likely contributed to the statistical significance of models. The ROCkit software assumes that scores for positive and negative cases follow a normal distribution, emphasizing consistent differences in predictions between the two models. Consequently, it can identify statistically significant differences even when overall AUC scores are very similar. Hence, the study's 167 mRS cases for 3-month outcomes offered sufficient statistical strength to identify small differences in AUC.

This study is the first in this project to develop a quantitative imaging biomarker based on regional volumetric analysis for ischemic stroke patients. The primary results focus on predicting mRS scores at discharge and assessing long-term outcomes at 3 months, using a combination of volumetric and clinical features. In future studies, the dataset will be expanded, and further analysis will be conducted by generating additional volume-based features and exploring various feature subsets. For example, future work will investigate features such as (i) mapped volumes relative to total ASPECTS and PC-ASPECTS volumes combined with transformed clinical features, and (ii) mapped volumes relative to edema volume alongside clinical features. This extended analysis aims to explore the potential of multivolumetric features as imaging biomarkers for AIS patients.

In summary, this pilot study demonstrated that volumetric features of ASPECTS and PC-ASPECTS regions relative to overall brain volume are highly effective and relevant in-

dicators of future functional outcomes in stroke-affected patients.

CHAPTER 6

Conclusion and Futurework

Advances in imaging technology have given doctors and radiologists access to large amounts of data, greatly improving diagnostic accuracy. However, analyzing such vast datasets has become increasingly complex, leading to a growing need for a "second opinion." This has driven the development of AI-integrated CAD tools for detection and diagnosis in various medical fields. Researchers have created several AI-CAD models using manual and automated feature extraction techniques, making these tools an essential part of modern healthcare. Beyond detection, AI-CAD models significantly improve disease prediction and outcome forecasting, helping design better treatments for patients while reducing unnecessary procedures and costs.

Stroke is a major global health concern, marked by high rates of mortality and morbidity. Survivors often face significant challenges, including functional impairment, emotional side effects, and recurrence, all of which diminish their quality of life. These issues underscore the need for improved stroke management and prevention strategies. With that objective, this dissertation aims to develop a predictive model for stroke prognosis, ultimately enhancing patient outcomes through better hospital treatment planning, personalized follow-up care,

and efficient resource allocation. This dissertation presents three studies: the first two focus on aSAH patients, while the third addresses AIS patients. All studies utilized data provided by a neurologist, whose clinical expertise was instrumental in identifying and addressing persistent gaps in current stroke care practices.

The first study in this dissertation examined the effectiveness of deep-learning-based classification models in developing radiographic imaging biomarkers to predict both short-term clinical events and long-term outcomes for patients with aSAH. Addressing the limitations of traditional assessments—which tend to be subjective and subject to significant inter-rater variability—this study proposed an attention-integrated deep transfer learning architecture. This approach aimed to better leverage imaging biomarkers for predicting stroke outcomes, a relatively underexplored area. A novel imaging biomarker was introduced by utilizing blood volume in the cisternal space of the brain, a visible marker on CT scans, and a recognized risk factor. The study evaluated three outcomes: two short-term and one long-term, to demonstrate the prediction model's potential. The model's high sensitivity (0.88), AUC score from the ROC curve (0.78 ± 0.14), and F1 score (0.82) underscore its strong discriminative ability to identify patients at risk of adverse outcomes, thus supporting more timely interventions by medical professionals, resource allocation, and planning. The study demonstrated impressive results using an attention-integrated deep framework to extract imaging biomarkers; however, the approach remains holistic from a clinical standpoint. Future research could explore additional radiomic features beyond sulcal volume and develop fusion models that integrate clinical, radiomic, and deep features to provide more explainable results.

The second study of the dissertation aimed at addressing the limitations of the first study by expanding the exploration of imaging biomarkers with an ensembled hybrid approach in developing the prognostic model and did a comparative analysis of prediction models with different approaches, including radiomic, deep transfer learning, and hybrid methods. The conventional radiomic feature extraction involves manually crafted features to obtain quantitative information with available domain knowledge, while the DL method provides a more comprehensive and holistic approach capable of identifying intricate patterns in

the input data. Recognizing the unique contributions of each technique, features extracted from both methods were integrated and assessed with multiple ML-based models to present the optimized hybrid prediction model. In this study, the assessments were extended and included a total of eight outcomes: four short-term and four long-term clinical outcomes. Radiological scans were curated and processed by segmenting and labeling multiple regions of the brain for further analysis. Feature engineering techniques were applied to extract radiomic features based on the sulcal volume of the brain to other brain regions, which is an established prognostic tool. Then deep features extracted from multiple deep transfer learning models were integrated.

The features from the four deep models from the first study were fused with radiomic features, respectively, to create the hybrid model. One model combines radiomic and all deep features, which gives five hybrid models. PCA was used to reduce the feature set, while SMOTE addressed class imbalance during LOO-CV. Five different machine learning models were adopted to analyze these five hybrid models, which were evaluated with an AUC score from the ROC curve, sensitivity, specificity, accuracy, precision, and F1 score. To find the optimized prognostic model for each outcome, a thorough comparative analysis was done, which showed that in most outcomes, the hybrid model with ensemble features performed the best.

The findings of the study revealed that the hybrid approach of combining traditional ML and DL architectures outperformed the performance of solely radiomic or DL-based approaches using the same dataset. The higher AUC scores indicated the model's strong capability in distinguishing between negative and positive outcomes in patients, while high accuracy confirmed the correctness of the classifier-based prediction model. Overall, this study demonstrates the potential of hybrid imaging biomarkers to improve predictive accuracy and efficacy in clinical practice, contributing to better patient outcomes and resource allocation.

Moreover, both my first and second studies highlight the importance of scan timing in predicting outcomes. Admission scans proved to be more effective for assessing short-

term clinical events, as early imaging can provide valuable information for post-intervention planning and decisions to manage early brain injury (EBI). In contrast, discharge scans offered greater insight into predicting long-term functional outcomes, revealing the impact of treatment on patients and giving clinicians essential information to tailor recovery plans upon discharge. These findings support the clinical approach of using initial scans to monitor early developments, while follow-up scans help assess treatment effectiveness and guide discharge planning. Together, the two studies of aSAH cases underscore the role of imaging timing in outcome prediction, providing valuable insights for optimizing patient management and recovery strategies.

To expand the scope of the core objective of this dissertation, the third study focused on developing a composite tool to predict functional outcomes for ischemic stroke patients. This study reported the pilot project in which a composite tool was created, utilizing volumetric analysis of ischemic and total brain volumes. Radiological features from imaging scans across various modalities were integrated with demographic and clinical characteristics to enhance predictive power. Unlike traditional grading-based methods for post-AIS treatment outcome prediction, this study analyzed volumes in ASPECTS and PC-ASPECTS regions, ischemic areas, and overall brain regions. ASPECTS/PC-ASPECTS regions were annotated and mapped with doctor-marked ischemic volumes to extract relevant features. Additionally, transformed features were generated by calculating proportional volumes of mapped regions relative to the total ischemic volume, as well as synthesized volumes through summation and proportional analysis of relevant regions. The study assessed functional outcomes using the Modified Rankin Scale (mRS) at discharge and a 3-month follow-up.

To identify an optimized model, four popular machine learning models were tested and evaluated across several performance metrics. Feature engineering techniques, including Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), LASSO, and Elastic Net (EN), were applied to select and refine relevant features. With a high AUC score of 0.94 ± 0.02, this preliminary study demonstrated the effectiveness of volumetric features in predicting functional outcomes. Among the four models and five feature engineering

techniques, Elastic Net and RFE yielded the best performance with SVM and XGB models, respectively. Additionally, the transformed feature set slightly outperformed the raw features across all feature selection techniques, underscoring the potential for enhanced accuracy with engineered features. The pilot study demonstrated the potential of volumetric features in prognosis assessment for ischemic stroke patients. While this study utilized a selection of volumetric features, additional features could be developed and tested to expand the project's scope. Future work could involve generating more transformed features by incorporating a broader range of volume-based metrics. This study assessed 187 cases, but including more data will enhance model learning and generalizability.

Despite significant progress in CAD research, challenges remain in making these systems clinically applicable. While advanced deep learning systems excel in predicting future outcomes, they have some limitations. For example, the models often tested on small datasets with augmentation techniques applied to increase datasets and their "black-box" feature extraction approach makes it difficult to gain the trust of clinicians. To address this, more researches need to be done to develop explainable and visually interactive CAD systems, enabling clinical researchers to confidently use these tools in practical settings.

Building on my Ph.D. experience, I aim to continue researching advanced techniques to improve stroke prognostic models and enhance their acceptance and usability in clinical practice. In the first two studies, aSAH cases were sourced from a single institution. Future research could benefit from incorporating data from multiple institutions to capture a broader range of features, thereby improving model robustness. Exploring hybrid models that integrate clinical data with multiple radiomic and deep features could also offer significant advancements. Additionally, a larger dataset would allow for the investigation of advanced deep architectures, such as Vision Transformers (ViT). For AIS stroke patients, future work could focus on generating additional volumetric features by mapping volumes to ASPECTS and PC-ASPECTS regions and combining these with clinical features. A comprehensive feature set could be created by integrating ischemic region volumes with ASPECTS/PC-ASPECTS volumes and clinical data. Further analysis could identify which

combinations of features serve as the most effective prognostic biomarkers. I also plan to leverage the deep learning framework developed for the aSAH study to explore additional prognostic biomarkers for AIS patients. Combining these with clinical and radiomic features would provide a more comprehensive understanding, aiding the development of more reliable prognostic models for stroke patients and ultimately enhancing patient care.

In summary, my Ph.D. research provided an invaluable opportunity to explore, investigate, and contribute to the development and testing of innovative CAD schemes. These efforts culminated in presenting a robust prognostic model aimed at improving outcomes for stroke-affected patients, ultimately enhancing clinical decision-making and patient care. Building on my prior research and the progress made during my Ph.D. studies, I aim to address the challenges associated with the proposed prediction models and develop more robust and reliable approaches. By focusing on integrating CAD systems with deep learning and hybrid models, my work has the potential to significantly influence the design of stroke patient care. The outcomes of this research are expected to contribute to improving patient outcomes and expanding the practical application of these systems in healthcare. Additionally, the knowledge and achievements gained during my Ph.D. will provide a solid foundation for my ongoing academic endeavors, professional development, and future career, enabling me to drive further innovation in this vital field.

Bibliography

[1] Hiroshi Fujita. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first, 3 2020.

[2] Intisar Rizwan I Haque and Jeremiah Neubert. Deep learning approaches to biomedical image segmentation, 1 2020.

[3] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, 32(4):582–596, 8 2019.

[4] Abdulrahman Abbas Mukhlif, Belal Al-Khateeb, and Mazin Abed Mohammed. An extensive review of state-of-the-art transfer learning techniques used in medical imaging: Open issues and challenges, 1 2022.

[5] Dong Keun Kim, Eung Y. Kim, Kun H. Yang, Chung Ki Lee, and Sun K. Yoo. A mobile tele-radiology imaging system with JPEG2000 for an emergency care. *Journal of Digital Imaging*, 24(4):709–718, 8 2011.

[6] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. 8 2016.

[7] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional Block Attention Module. 7 2018.

[8] Kunio Doi. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. Technical report.

[9] Ronald A. Castellino. Computer aided detection (CAD): An overview, 2005.

[10] Eman Justaniah, Areej Alhothali, and Dr Ghadah Aldabbagh. Mammogram Segmentation Techniques: A Review. Technical Report 5.

[11] Marilyn J. Morton, Dana H. Whaley, Kathleen R. Brandt, and Kimberly K. Amrami. Screening mammograms: Interpretation with computer-aided detection-prospective evaluation, 5 2006.

[12] Yuan Peng Zhang, Xin Yun Zhang, Yu Ting Cheng, Bing Li, Xin Zhi Teng, Jiang Zhang, Saikit Lam, Ta Zhou, Zong Rui Ma, Jia Bao Sheng, Victor C.W. Tam, Shara W.Y. Lee, Hong Ge, and Jing Cai. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling, 12 2023.

[13] Luís Pinto-Coelho. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications, 12 2023.

[14] Most Nilufa Yeasmin, Md Al Amin, Tasmim Jamal Joti, Zeyar Aung, and Mohammad Abdul Azim. Advances of AI in Image-Based Computer-Aided Diagnosis: A Review. 2023.

[15] Mengfang Li, Yuanyuan Jiang, Yanzhou Zhang, and Haisheng Zhu. Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, 11, 2023.

[16] Reabal Najjar. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging, 9 2023.

[17] Vishwa S. Parekh and Michael A. Jacobs. Deep learning and radiomics in precision medicine, 3 2019.

[18] Laura M. Winchester, Eric L. Harshfield, Liu Shi, Aman Preet Badhwar, Ahmad Al Khleifat, Natasha Clarke, Amir Dehsarvi, Imre Lengyel, Ilianna Lourida, Christopher R. Madan, Sarah J. Marzi, Petroula Proitsi, Anto P. Rajkumar, Timothy Rittman, Edina Silajdžić, Stefano Tamburin, Janice M. Ranson, and David J. Llewellyn. Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia, 12 2023.

[19] Gehad A. Saleh, Nihal M. Batouty, Abdelrahman Gamal, Ahmed Elnakib, Omar Hamdy, Ahmed Sharafeldeen, Ali Mahmoud, Mohammed Ghazal, Jawad Yousaf, Marah Alhalabi, Amal AbouEleneen, Ahmed Elsaid Tolba, Samir Elmougy, Sohail Contractor, and Ayman El-Baz. Impact of Imaging Biomarkers and AI on Breast Cancer Management: A Brief Review, 11 2023.

[20] Minliang He, Xuming Wang, and Yijun Zhao. A calibrated deep learning ensemble for abnormality detection in musculoskeletal radiographs. *Scientific reports*, 11(1):9097, 4 2021.

[21] Alanna Vial, David Stirling, Matthew Field, Montserrat Ros, Christian Ritz, Martin Carolan, Lois Holloway, and Alexis A. Miller. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review, 6 2018.

[22] Farah Shamout, Tingting Zhu, and David A. Clifton. Machine Learning for Clinical Outcome Prediction. *IEEE Reviews in Biomedical Engineering*, 14:116–126, 2021.

[23] Nenad Tomašev, Xavier Glorot, Jack W. Rae, Michal Zielinski, Harry Askham, Andre Saraiva, Anne Mottram, Clemens Meyer, Suman Ravuri, Ivan Protsyuk, Alistair Connell, Cían O. Hughes, Alan Karthikesalingam, Julien Cornebise, Hugh Montgomery, Geraint Rees, Chris Laing, Clifton R. Baker, Kelly Peterson, Ruth Reeves, Demis Hassabis, Dominic King, Mustafa Suleyman, Trevor Back, Christopher Nielson, Joseph R. Ledsam, and Shakir Mohamed. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572(7767):116–119, 8 2019.

[24] Min Chen, Yixue Hao, Kai Hwang, Lin Wang, and Lu Wang. Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access*, 5:8869–8879, 2017.

[25] Zijian Zhu, Lai Jiang, and Xianting Ding. Advancing Breast Cancer Heterogeneity Analysis: Insights from Genomics, Transcriptomics and Proteomics at Bulk and Single-Cell Levels, 8 2023.

[26] Fergus Davnall, Connie S.P. Yip, Gunnar Ljungqvist, Mariyah Selmi, Francesca Ng, Bal Sanghera, Balaji Ganeshan, Kenneth A. Miles, Gary J. Cook, and Vicky Goh. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice?, 12 2012.

[27] Wenya Linda Bi, Ahmed Hosny, Matthew B. Schabath, Maryellen L. Giger, Nicolai J. Birkbak, Alireza Mehrtash, Tavis Allison, Omar Arnaout, Christopher Abbosh, Ian F. Dunn, Raymond H. Mak, Rulla M. Tamimi, Clare M. Tempany, Charles Swanton, Udo Hoffmann, Lawrence H. Schwartz, Robert J. Gillies, Raymond Y. Huang, and Hugo J. W. L. Aerts. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: A Cancer Journal for Clinicians*, 69(2):127–157, 3 2019.

[28] Nayla Nasr, Beáta Soltész, János Sándor, Róza Adány, and Szilvia Fiatal. Prognostic Modelling Studies of Coronary Heart Disease—A Systematic Review of Conventional and Genetic Risk Factor Studies, 9 2022.

[29] Karel G.M. Moons, Andre Pascal Kengne, Mark Woodward, Patrick Royston, Yvonne Vergouwe, Douglas G. Altman, and Diederick E. Grobbee. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker, 5 2012.

[30] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. Prediction of Heart Disease Using a Combination of Machine

Learning and Deep Learning. *Computational Intelligence and Neuroscience*, 2021, 2021.

[31] J. E. Soun, D. S. Chow, M. Nagamine, R. S. Takhtawala, C. G. Filippi, W. Yu, and P. D. Chang. Artificial intelligence and acute stroke imaging, 1 2021.

[32] Ahmad A. Abujaber, Ibraheem M. Alkhawaldeh, Yahia Imam, Abdulqadir J. Nashwan, Naveed Akhtar, Ahmed Own, Ahmad S. Tarawneh, and Ahmad B. Hassanat. Predicting 90-day prognosis for patients with stroke: a machine learning approach. *Frontiers in Neurology*, 14, 2023.

[33] Saeed Shurrab, Alejandro Guerra-Manzanares, Amani Magid, Bartlomiej Piechowski-Jozwiak, S. Farokh Atashzar, and Farah E. Shamout. Multimodal Machine Learning for Stroke Prognosis and Diagnosis: A Systematic Review. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[34] Rüdigger Von Kummer. Time is brain: Fact or fiction, 3 2019.

[35] Isha R. Chavva, Anna L. Crawford, Mercy H. Mazurek, Matthew M. Yuen, Anjali M. Prabhat, Sam Payabvash, Gordon Sze, Guido J. Falcone, Charles C. Matouk, Adam de Havenon, Jennifer A. Kim, Richa Sharma, Steven J. Schiff, Matthew S. Rosen, Jayashree Kalpathy-Cramer, Juan E. Iglesias Gonzalez, W. Taylor Kimberly, and Kevin N. Sheth. Deep Learning Applications for Acute Stroke Management, 10 2022.

[36] Yue Zhang, Shijie Liu, Chunlai Li, and Jianyu Wang. Application of Deep Learning Method on Ischemic Stroke Lesion Segmentation, 2 2022.

[37] Asit Subudhi, Pratyusa Dash, Manoranjan Mohapatra, Ru-San Tan, U Rajendra Acharya, and Sukanta Sabut. Application of Machine Learning Techniques for Characterization of Ischemic Stroke with MRI Images: A Review. *Diagnostics 2022*, 12:2535, 2535.

[38] Nathan A. Shlobin, Ammad A. Baig, Muhammad Waqas, Tatsat R. Patel, Rimal H. Dossani, Megan Wilson, Justin M. Cappuzzo, Adnan H. Siddiqui, Vincent M. Tutino, and Elad I. Levy. Artificial Intelligence for Large-Vessel Occlusion Stroke: A Systematic Review, 3 2022.

[39] Christian Maegerlein, Johanna Fischer, Sebastian Mönch, Maria Berndt, Silke Wunderlich, Christian L. Seifert, Manuel Lehm, Tobias Boeckh-Behrens, Claus Zimmer, and Benjamin Friedrich. Automated calculation of the Alberta stroke program early CT score: Feasibility and reliability. *Radiology*, 291(1):141–148, 4 2019.

[40] Jaehak Yu, Sejin Park, Soon Hyun Kwon, Chee Meng Benjamin Ho, Cheol Sig Pyo, and Hansung Lee. AI-based stroke disease prediction system using real-time electromyography signals. *Applied Sciences (Switzerland)*, 10(19), 10 2020.

[41] Kim Mouridsen, Patrick Thurner, and Greg Zaharchuk. Artificial Intelligence Applications in Stroke, 8 2020.

[42] Nick M. Murray, Mathias Unberath, Gregory D. Hager, and Ferdinand K. Hui. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: A systematic review, 2 2020.

[43] Vida Abedi, Venkatesh Avula, Seyed Mostafa Razavi, Shreya Bavishi, Durgesh Chaudhary, Shima Shahjouei, Ming Wang, Christoph J. Griessenauer, Jiang Li, and Ramin Zand. Predicting short and long-term mortality after acute ischemic stroke using EHR. *Journal of the Neurological Sciences*, 427:117560, 8 2021.

[44] Nitika Goenka and Shamik Tiwari. Deep learning for Alzheimer prediction using brain biomarkers. *Artificial Intelligence Review*, 54(7):4827–4871, 10 2021.

[45] Gonçalo Oliveira, Ana Catarina Fonseca, José Ferro, and Arlindo L Oliveira. Deep learning-based extraction of biomarkers for the prediction of the functional outcome of ischemic stroke patients. *Diagnostics*, 13(24):3604, 2023.

[46] Wessel E. Van Der Steen, H. A. Marquering, L. A. Ramos, R. Van Den Berg, B. A. Coert, A. M.M. Boers, M. D.I. Vergouwen, G. J.E. Rinkel, B. K. Velthuis, Y. B.W.E.M. Roos, C. B.L.M. Majoie, W. P. Vandertop, and D. Verbaan. Prediction of outcome using quantified blood volume in aneurysmal SAH. *American Journal of Neuroradiology*, 41(6):1015–1021, 6 2020.

[47] Sávio Batista, Jhon E. Bocanegra-Becerra, Bernardo Claassen, Felipe Rubião, Nicollas Nunes Rabelo, Eberval Gadelha Figueiredo, and Dan Zimelewicz Oberman. Biomarkers in aneurysmal subarachnoid hemorrhage: A short review, 7 2023.

[48] Stefan Yu Bögli, Ihsane Olakorede, Michael Veldeman, Erta Beqiri, Miriam Weiss, Gerrit Alexander Schubert, Jan Folkard Willms, Emanuela Keller, and Peter Smielewski. Predicting outcome after aneurysmal subarachnoid hemorrhage by exploitation of signal complexity: a prospective two-center cohort study. *Critical Care*, 28(1), 12 2024.

[49] I. A. Zijlstra, C. S. Gathier, A. M. Boers, H. A. Marquering, A. J. Slooter, B. K. Velthuis, B. A. Coert, D. Verbaan, R. Van Den Berg, G. J. Rinkel, and C. B. Majoie. Association of automatically quantified total blood volume after aneurysmal subarachnoid hemorrhage with delayed cerebral ischemia. In *American Journal of Neuroradiology*, volume 37, pages 1588–1593. American Society of Neuroradiology, 9 2016.

[50] Claire Delpirou Nouh, Bappaditya Ray, Chao Xu, Bin Zheng, Gopichand Danala, Ahmed Koriesh, Kimberly Hollabaugh, David Gordon, and Evgeny V. Sidorov. Quantitative Analysis of Stress-Induced Hyperglycemia and Intracranial Blood Volumes for Predicting Mortality After Intracerebral Hemorrhage. *Translational Stroke Research*, 13(4):595–603, 8 2022.

[51] Kyle P. O'Connor, Mausaminben Y. Hathidara, Gopichandh Danala, Chao Xu, Tressie M. McCoy, Evgeny V. Sidorov, Bin Zheng, Bradley N. Bohnstedt, and Bappaditya Ray. Predicting Clinical Outcome After Mechanical Thrombectomy: The GADIS

(Gender, Age, Diabetes Mellitus History, Infarct Volume, and Sex) Score. *World Neurosurgery*, 134:e1130–e1142, 2 2020.

[52] Philip A. Barber, Andrew M. Demchuk, Jinjin Zhang, and Alastair M. Buchan. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet*, 355(9216):1670–1674, 5 2000.

[53] Shelagh B. Coutts, Michael H. Lev, Michael Eliasziw, Luca Roccatagliata, Michael D. Hill, Lee H. Schwamm, J. H.Warwick Pexman, Walter J. Koroshetz, Mark E. Hudon, Alastair M. Buchan, R. Gilberto Gonzalez, and Andrew M. Demchuk. ASPECTS on CTA source images versus unenhanced CT: Added value in predicting final infarct extent and clinical outcome. In *Stroke*, volume 35, pages 2472–2476, 11 2004.

[54] Gianluca Brugnara, Peter Mihalicz, Christian Herweh, Silvia Schönenberger, Jan Purrucker, Simon Nagel, Peter Arthur Ringleb, Martin Bendszus, Markus A. Möhlenbruch, and Ulf Neuberger. Clinical value of automated volumetric quantification of early ischemic tissue changes on non-contrast CT. *Journal of NeuroInterventional Surgery*, 15(e2):E178–E183, 11 2023.

[55] S. M. Seyedsaadat, A. A. Neuhaus, P. J. Nicholson, E. C. Polley, C. A. Hilditch, D. C. Mihal, T. Krings, J. Benson, I. Mark, D. F. Kallmes, W. Brinjikji, and J. D. Schaafsma. Differential contribution of ASPECTS regions to clinical outcome after thrombectomy for acute ischemic stroke. *American Journal of Neuroradiology*, 42(6):1104–1108, 6 2021.

[56] Heang Ping Chan, Lubomir M. Hadjiiski, and Ravi K. Samala. Computer-aided diagnosis in the era of deep learning. In *Medical Physics*, volume 47, pages e218–e227. John Wiley and Sons Ltd, 6 2020.

[57] Heang-Ping Chan, Kunio Doi, Carl J Vybrony, Robert A Schmidt, Charles E Metz, Kwok Leung Lam, Toshihiro Ogura, Yuzheng Wu, and Heber MacMahon. Improve-

ment in Radiologists' Detection of Clustered Microcalcifications on Mammograms: The Potential of Computer-Aided Diagnosis. *Investigative Radiology*, 25:1102–1110, 1990.

[58] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning, 5 2015.

[59] Kenji Suzuki. Overview of deep learning in medical imaging, 9 2017.

[60] Scott B. Hu, Deborah J.L. Wong, Aditi Correa, Ning Li, and Jane C. Deng. Prediction of clinical deterioration in hospitalized adult patients with hematologic malignancies using a neural network model. *PLoS ONE*, 11(8), 8 2016.

[61] Farah Shamout, Tingting Zhu, and David A Clifton. Machine learning for clinical outcome prediction. *IEEE reviews in Biomedical Engineering*, 14:116–126, 2020.

[62] Isaac S. Chan and Geoffrey S. Ginsburg. Personalized medicine: Progress and promise. *Annual Review of Genomics and Human Genetics*, 12:217–244, 7 2011.

[63] James Meehan, Mark Gray, Carlos Martínez-Pérez, Charlene Kay, Lisa Y. Pang, Jennifer A. Fraser, Amy V. Poole, Ian H. Kunkler, Simon P. Langdon, David Argyle, and Arran K. Turnbull. Precision Medicine and the Role of Biomarkers of Radiotherapy Response in Breast Cancer, 4 2020.

[64] Andrzej Kloczkowski, Cristina Peña, Qasem Al-Tashi, Maliazurina B Saad, Amgad Muneer, Rizwan Qureshi, Seyedali Mirjalili, Ajay Sheshadri, Xiuning Le, Natalie I Vokes, Jianjun Zhang, and Jia Wu. Machine Learning Models for the Identification of Prognostic and Predictive Cancer Biomarkers: A Systematic Review. *Int. J. Mol. Sci*, 2023:7781, 2023.

[65] Amy P. Abernethy, Lynn M. Etheredge, Patricia A. Ganz, Paul Wallace, Robert R. German, Chalapathy Neti, Peter B. Bach, and Sharon B. Murphy. Rapid-learning system for cancer care. *Journal of Clinical Oncology*, 28(27):4268–4274, 9 2010.

[66] Karla V. Ballman. Biomarker: Predictive or prognostic? *Journal of Clinical Oncology*, 33(33):3968–3971, 11 2015.

[67] Ralph Th. TheranosTics and precision medicine special feaTure: review arTicle a review on radiomics and the future of theranostics for patient selection in precision medicine. Technical report, 2018.

[68] Md Sohanur Rahman, Khandaker Reajul Islam, Johayra Prithula, Jaya Kumar, Mufti Mahmud, Mohammed Fasihul Alam, Mamun Bin Ibne Reaz, Abdulrahman Alqahtani, and Muhammad E.H. Chowdhury. Machine learning-based prognostic model for 30-day mortality prediction in Sepsis-3. *BMC Medical Informatics and Decision Making*, 24(1), 12 2024.

[69] Rahila Sardar, Arun Sharma, and Dinesh Gupta. Machine Learning Assisted Prediction of Prognostic Biomarkers Associated With COVID-19, Using Clinical and Proteomics Data. *Frontiers in Genetics*, 12, 5 2021.

[70] Freddie Márquez and Michael A. Yassa. Neuroimaging Biomarkers for Alzheimer's Disease, 6 2019.

[71] Marcel Koenigkam Santos, José Raniery Ferreira Júnior, Danilo Tadao Wada, Ariane Priscilla Magalhães Tenório, Marcello Henrique Nogueira Barbosa, and Paulo Mazzoncini De Azevedo Marques. Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: Advances in imaging towards to precision medicine. *Radiologia Brasileira*, 52(6):387–396, 11 2019.

[72] Shiju Yan, Wei Qian, Yubao Guan, and Bin Zheng. Improving lung cancer prognosis assessment by incorporating synthetic minority oversampling technique and score fusion method. *Medical Physics*, 43(6):2694–2703, 6 2016.

[73] Norma Latif Fitriyani, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension. *IEEE Access*, 7:144777–144789, 2019.

[74] Isabel Fragata, Marta Alves, Ana Luísa Papoila, Patrícia Ferreira, Ana Paiva Nunes, Nuno Canto Moreira, and Patrícia Canhão. Prediction of clinical outcome in subacute

subarachnoid hemorrhage using diffusion tensor imaging. *Journal of Neurosurgery*, 130(2):550–558, 2 2019.

[75] Ahmed Esmael, Mohammed Elsherief, and Khaled Eltoukhy. Predictive Value of the Alberta Stroke Program Early CT Score (ASPECTS) in the Outcome of the Acute Ischemic Stroke and Its Correlation with Stroke Subtypes, NIHSS, and Cognitive Impairment. *Stroke Research and Treatment*, 2021, 2021.

[76] Shujun Huang, C. A.I. Nianguang, Pedro Penzuti Pacheco, Shavira Narandes, Yang Wang, and X. U. Wayne. Applications of support vector machine (SVM) learning in cancer genomics, 1 2018.

[77] Chip M. Lynch, Behnaz Abdollahi, Joshua D. Fuqua, Alexandra R. de Carlo, James A. Bartholomai, Rayeanne N. Balgemann, Victor H. van Berkel, and Hermann B. Frieboes. Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International Journal of Medical Informatics*, 108:1–8, 12 2017.

[78] Luca Steardo, Elvira Anna Carbone, Renato de Filippis, Claudia Pisanu, Cristina Segura-Garcia, Alessio Squassina, Pasquale De Fazio, and Luca Steardo. Application of support vector machine on fmri data as biomarkers in schizophrenia diagnosis: A systematic review. *Frontiers in Psychiatry*, 11, 2020.

[79] Maren E. Shipe, Stephen A. Deppen, Farhood Farjah, and Eric L. Grogan. Developing prediction models for clinical use using logistic regression: An overview, 2019.

[80] Martin Krämer, Maja Ingwersen, Ulf Teichgräber, and Felix Güttler. Added value of chest CT in a machine learning-based prediction model to rule out COVID-19 before inpatient admission: A retrospective university network study. *European Journal of Radiology*, 163, 6 2023.

[81] Ji Hwan Park, Han Eol Cho, Jong Hun Kim, Melanie M. Wall, Yaakov Stern, Hyunsun Lim, Shinjae Yoo, Hyoung Seop Kim, and Jiook Cha. Machine learning prediction

of incidence of Alzheimer's disease using large-scale administrative health data. *npj Digital Medicine*, 3(1), 12 2020.

[82] Shuja Mirza, Sonu Mittal, and Majid Zaman. Decision Support Predictive model for prognosis of diabetes using SMOTE and Decision tree. Technical Report 11, 2018.

[83] Kazuya Matsuo, Hideo Aihara, Tomoaki Nakai, Akitsugu Morishita, Yoshiki Tohma, and Eiji Kohmura. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. *Journal of Neurotrauma*, 37(1):202–210, 1 2020.

[84] Cuong Nguyen, Yong Wang, and Ha Nam Nguyen. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. *Journal of Biomedical Science and Engineering*, 06(05):551–560, 2013.

[85] Luming Zhang, Tao Huang, Fengshuo Xu, Shaojin Li, Shuai Zheng, Jun Lyu, and Haiyan Yin. Prediction of prognosis in elderly patients with sepsis based on machine learning (random survival forest). *BMC Emergency Medicine*, 22(1), 12 2022.

[86] Ravinder Ahuja, Subhash C. Sharma, and Maaruf Ali. A diabetic disease prediction model based on classification algorithms. *Annals of Emerging Technologies in Computing*, 3(3):44–52, 7 2019.

[87] Surbhi Gupta and Manoj Kumar. Prostate cancer prognosis using multi-layer perceptron and class balancing techniques. In *ACM International Conference Proceeding Series*, pages 1–6. Association for Computing Machinery, 8 2021.

[88] Rui Wen, Miaoran Wang, Wei Bian, Haoyue Zhu, Ying Xiao, Qian He, Yu Wang, Xiaoqing Liu, Yangdi Shi, Zhe Hong, and Bing Xu. Machine learning-based prediction of symptomatic intracerebral hemorrhage after intravenous thrombolysis for stroke: a large multicenter study. *Frontiers in Neurology*, 14, 2023.

[89] Siteng Chen, Liren Jiang, Feng Gao, Encheng Zhang, Tao Wang, Ning Zhang, Xiang Wang, and Junhua Zheng. Machine learning-based pathomics signature could act as

a novel prognostic marker for patients with clear cell renal cell carcinoma. *British Journal of Cancer*, 126(5):771–777, 3 2022.

[90] Rashmi Siddalingappa and Sekar Kanagaraj. K-nearest-neighbor algorithm to predict the survival time and classification of various stages of oral cancer: a machine learning approach. *F1000Research*, 11:70, 1 2022.

[91] Ghulam Muhammad, Saad Naveed, Lubna Nadeem, Tariq Mahmood, Amjad R. Khan, Yasar Amin, and Saeed Ali Omer Bahaj. Enhancing Prognosis Accuracy for Ischemic Cardiovascular Disease Using K Nearest Neighbor Algorithm: A Robust Approach. *IEEE Access*, 11:97879–97895, 2023.

[92] Zhiqiang Yan, Jiang Wang, Qiufeng Dong, Lian Zhu, Wei Lin, and Xiaofan Jiang. XGBoost algorithm and logistic regression to predict the postoperative 5-year outcome in patients with glioma. *Annals of Translational Medicine*, 10(16):860–860, 8 2022.

[93] Ruoran Wang, Jing Zhang, Baoyin Shan, Min He, and Jianguo Xu. XGBoost Machine Learning Algorithm for Prediction of Outcome in Aneurysmal Subarachnoid Hemorrhage. *Neuropsychiatric Disease and Treatment*, 18:659–667, 2022.

[94] Lei Cai, Jingyang Gao, and Di Zhao. A review of the application of deep learning in medical image classification and segmentation. *Annals of Translational Medicine*, 8(11):713–713, 6 2020.

[95] Niall O' Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. Deep Learning vs. Traditional Computer Vision. Technical report.

[96] Sadia Arooj, Saif ur Rehman, Azhar Imran, Abdullah Almuhaimeed, A. Khuzaim Alzahrani, and Abdulkareem Alzahrani. A Deep Convolutional Neural Network for the Early Detection of Heart Disease. *Biomedicines*, 10(11), 11 2022.

[97] Ahmad Waleed Salehi, Preety Baglat, Brij Bhushan Sharma, Ankita Upadhya, and Gaurav Gupta. *A CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease using MRI*. IEEE, 2020.

[98] Shubham Dwivedi, Tripti Goel, Rahul Sharma, and R. Murugan. Structural MRI based Alzheimer's Disease prognosis using 3D Convolutional Neural Network and Support Vector Machine. In *2021 Advanced Communication Technologies and Signal Processing, ACTS 2021*. Institute of Electrical and Electronics Engineers Inc., 2021.

[99] S Zhang, R Hou, X Cai, J Liu, W Yu, Q Zhang, J Guo, Z Zhu, W Feng, X Fu, Y Fu, L Qian, and Z Ding. Clinical Efficacy of Postoperative Radiotherapy for High-Risk Locoregional Recurrence in Patients with Completely Resected Pathologic Stage IIIA(N2) Non-Small Cell Lung Cancer: A Propensity Score-Matched Analysis A 3D CNN Model Using Deep Learning Based on Chest CT for Predicting Disease-Free Survival of Patients with Early-Stage Lung Cancer Receiving Surgery and SBRT. Technical report.

[100] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. 9 2014.

[101] Morteza Heidari, Seyedehnafiseh Mirniaharikandehei, Abolfazl Zargari Khuzani, Gopichandh Danala, Yuchen Qiu, and Bin Zheng. Detecting COVID-19 infected pneumonia from x-ray images using a deep learning model with image preprocessing algorithm. page 28. SPIE-Intl Soc Optical Eng, 2 2021.

[102] Bing Yan Lim, Khin Wee Lai, Khairunnisa Haiskin, K. A.Saneera Hemantha Kulathilake, Zhi Chao Ong, Yan Chai Hum, Samiappan Dhanalakshmi, Xiang Wu, and Xiaowei Zuo. Deep Learning Model for Prediction of Progressive Mild Cognitive Impairment to Alzheimer's Disease Using Structural MRI. *Frontiers in Aging Neuroscience*, 14, 6 2022.

[103] Nanyan Zhu, Chen Liu, Xinyang Feng, Dipika Sikka, Sabrina Gjerswold-Selleck, Scott A. Small, and Jia Guo. Deep learning identifies neuroimaging signatures of

alzheimer's disease using structural and synthesized functional Mri data. In *Proceedings - International Symposium on Biomedical Imaging*, volume 2021-April, pages 216–220. IEEE Computer Society, 4 2021.

[104] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. 12 2015.

[105] Anees Abrol, Manish Bhattarai, Alex Fedorov, Yuhui Du, Sergey Plis, and Vince Calhoun. Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease. *Journal of Neuroscience Methods*, 339, 6 2020.

[106] Sashank Sridhar and Sowmya Sanagavarapu. Detection and Prognosis Evaluation of Diabetic Retinopathy using Ensemble Deep Convolutional Neural Networks. Technical report.

[107] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely Connected Convolutional Networks. Technical report.

[108] V. Gokula Krishnan, M. V.Vijaya Saradhi, S. Sai Kumar, G. Dhanalakshmi, P. Pushpa, and V. Vijayaraja. Hybrid Optimization based Feature Selection with DenseNet Model for Heart Disease Prediction. *International Journal of Electrical and Electronics Research*, 11(2):253–261, 2023.

[109] Adarsh Vulli, Parvathaneni Naga Srinivasu, Madipally Sai Krishna Sashank, Jana Shafi, Jaeyoung Choi, and Muhammad Fazal Ijaz. Fine-Tuned DenseNet-169 for Breast Cancer Metastasis Prediction Using FastAI and 1-Cycle Policy. *Sensors*, 22(8), 4 2022.

[110] Abeer Abdulaziz Alarfaj, Hanan A. Hosni Mahmoud, and Alaaeldin M. Hafez. A Deep Learning Model for Stroke Patients' Motor Function Prediction. *Applied Bionics and Biomechanics*, 2022, 2022.

[111] Evan Calabrese, Jeffrey D. Rudie, Andreas M. Rauschecker, Javier E. Villanueva-Meyer, Jennifer L. Clarke, David A. Solomon, and Soonmee Cha. Combining radiomics

and deep convolutional neural network features from preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma. *Neuro-Oncology Advances*, 4(1), 1 2022.

[112] Palak Mahajan, Shahadat Uddin, Farshid Hajati, and Mohammad Ali Moni. Ensemble Learning for Disease Prediction: A Review, 6 2023.

[113] K. Srinivas, R. Gagana Sri, K. Pravallika, K. Nishitha, and Subba Rao Polamuri. COVID-19 prediction based on hybrid Inception V3 with VGG16 using chest X-ray images. *Multimedia Tools and Applications*, 2023.

[114] J. Venkatesh, Anita Titus, B. Arthi, J. Jebastine, Subhani Shaik, and E. Sakthivel. Empirical Analysis of Hybrid Machine Learning based Earlier Stage Covid Disease Prediction using Lung CT Images. In *Proceedings of the 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2022*. Institute of Electrical and Electronics Engineers Inc., 2022.

[115] W. Han, L. Qin, C. Bay, X. Chen, K. H. Yu, N. Miskin, A. Li, X. Xu, and G. Young. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *American Journal of Neuroradiology*, 41(1):40–48, 2020.

[116] Mengtian Lu, Yaqi Wang, Jiaqiang Tian, and Haifeng Feng. Application of deep learning and radiomics in the prediction of hematoma expansion in intracerebral hemorrhage: a fully automated hybrid approach. *Diagnostic and Interventional Radiology*, 0(0):0–0, 4 2024.

[117] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 12 2014.

[118] Esraa Hassan, Mahmoud Y. Shams, Noha A. Hikal, and Samir Elmougy. The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study. *Multimedia Tools and Applications*, 82(11):16591–16633, 5 2023.

[119] Ahmet Demirkaya, Jiasi Chen, and Samet Oymak. Exploring the Role of Loss Functions in Multiclass Classification. In *2020 54th Annual Conference on Information Sciences and Systems, CISS 2020*. Institute of Electrical and Electronics Engineers Inc., 3 2020.

[120] Sukhpal Kaur, Himanshu Aggarwal, and Rinkle Rani. Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease. In *Machine Vision and Applications*, volume 31. Springer, 7 2020.

[121] Shui Hua Wang, Chaosheng Tang, Junding Sun, Jingyuan Yang, Chenxi Huang, Preetha Phillips, and Yu Dong Zhang. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Frontiers in Neuroscience*, 12(NOV), 11 2018.

[122] Hamid Moghaddasi, Bahareh Ahmadzadeh, Reza Rabiei, and Mohammad Farahbakhsh. Study on the Efficiency of a Multi-layer Perceptron Neural Network Based on the Number of Hidden Layers and Nodes for Diagnosing Coronary- Artery Disease. *Jentashapir Journal of Health Research*, 8(3), 6 2017.

[123] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1), 12 2019.

[124] Karl Weiss, Taghi M. Khoshgoftaar, and Ding Ding Wang. A survey of transfer learning. *Journal of Big Data*, 3(1), 12 2016.

[125] Noman Raza, Asma Naseer, Maria Tamoor, and Kashif Zafar. Alzheimer Disease Classification through Transfer Learning Approach. *Diagnostics*, 13(4), 2 2023.

[126] Jie Chen, Yingru Chen, Jianqiang Li, Jia Wang, Zijie Lin, and Asoke K Nandi. Stroke risk prediction with hybrid deep transfer learning framework. *IEEE Journal of Biomedical and Health Informatics*, 26(1):411–422, 2021.

[127] Davide Anguita, Luca Ghelardoni, Alessandro Ghio, Luca Oneto, and Sandro Ridella. *The 'K' in K-fold Cross Validation.*

[128] Chye Rou Shi and Robiah Adnan. Modified cross-validation as a method for estimating parameter. In *AIP Conference Proceedings*, volume 1635, pages 724–731. American Institute of Physics Inc., 2014.

[129] Bouchaib Cherradi, Oumaima Terrada, Asmae Ouhmida, Soufiane Hamida, Abdelhadi Raihani, and Omar Bouattane. Computer-Aided Diagnosis System for Early Prediction of Atherosclerosis using Machine Learning and K-fold cross-validation. In *2021 International Congress of Advanced Technology and Engineering, ICOTEN 2021*. Institute of Electrical and Electronics Engineers Inc., 7 2021.

[130] Dwi Normawati and Dewi Pramudi Ismi. K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining. *Signal and Image Processing Letters*, 1(2):23–35, 7 2019.

[131] Mobarakol Islam, Vibashan Vs, V Jeya, Maria Jose, Navodini Wijethilake, Uppal Utkarsh, and Hongliang Ren. Brain Tumor Segmentation and Survival Prediction using 3D Attention UNet. Technical report.

[132] Qingji Guan, Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng, and Yi Yang. Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. 1 2018.

[133] Željko Vujović. Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, 12(6):599–606, 2021.

[134] Thomas F. Monaghan, Syed N. Rahman, Christina W. Agudelo, Alan J. Wein, Jason M. Lazar, Karel Everaert, and Roger R. Dmochowski. Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value, 2021.

[135] Steven A. Hicks, Inga Strümke, Vajira Thambawita, Malek Hammou, Michael A. Riegler, Pål Halvorsen, and Sravanthi Parasa. On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, 12(1), 12 2022.

[136] Valery L. Feigin, Michael Brainin, Bo Norrving, Sheila Martins, Ralph L. Sacco, Werner Hacke, Marc Fisher, Jeyaraj Pandian, and Patrice Lindsay. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022, 1 2022.

[137] Thinni Nurul Rochmah, Indana Tri Rahmawati, Maznah Dahlui, Wasis Budiarto, and Nabilah Bilqis. Economic burden of stroke disease: A systematic review, 7 2021.

[138] Mira Katan and Andreas Luft. Global Burden of Stroke. *Seminars in Neurology*, 38(2):208–211, 4 2018.

[139] Ahmad Hassan, Saima Gulzar Ahmad, Ehsan Ullah Munir, Imtiaz Ali Khan, and Naeem Ramzan. Predictive modelling and identification of key risk factors for stroke using machine learning. *Scientific Reports*, 14(1), 12 2024.

[140] Emilia Salvadori, Gioele Papi, Greta Insalata, Valentina Rinnoci, Ida Donnini, Monica Martini, Catuscia Falsini, Bahia Hakiki, Annamaria Romoli, Carmen Barbato, Paola Polcaro, Francesca Casamorata, Claudio Macchi, Francesca Cecchi, and Anna Poggesi. Comparison between ischemic and hemorrhagic strokes in functional outcome at discharge from an intensive rehabilitation hospital. *Diagnostics*, 11(1), 1 2021.

[141] Mervyn D.I. Vergouwen and Gabriel J.E. Rinkel. Emergency Medical Management of Aneurysmal Subarachnoid Hemorrhage. *Neurocritical Care*, 39(1):51–58, 8 2023.

[142] Jane Y. Yuan, Yasheng Chen, Atul Kumar, Zach Zlepper, Keshav Jayaraman, Wint Y. Aung, Julian V. Clarke, Michelle Allen, Umeshkumar Athiraman, Joshua Osbun, Gregory J. Zipfel, and Rajat Dhar. Automated Quantification of Reduced Sulcal Volume Identifies Early Brain Injury After Aneurysmal Subarachnoid Hemorrhage. *Stroke*, 52(4):1380–1389, 4 2021.

[143] David C. Lauzier, Keshav Jayaraman, Jane Y. Yuan, Deepti Diwan, Ananth K. Vellimana, Joshua W. Osbun, Arindam R. Chatterjee, Umeshkumar Athiraman, Rajat Dhar, and Gregory J. Zipfel. Early Brain Injury After Subarachnoid Hemorrhage: Incidence and Mechanisms, 5 2023.

[144] Charles L. Francoeur and Stephan A. Mayer. Management of delayed cerebral ischemia after subarachnoid hemorrhage, 10 2016.

[145] Jennifer A Frontera, Jan Claassen, J Michael Schmidt, Katja E Wartenberg, Richard Temes, E Sander Connolly, R Loch Macdonald, and Stephan A Mayer. PREDICTION OF SYMPTOMATIC VASOSPASM AFTER SUBARACHNOID HEMORRHAGE: THE MODIFIED FISHER SCALE CLINICAL STUDIES. Technical report, 2006.

[146] Yu Chang Wang, Xiao Qiang Wang, Chang Wu Tan, Chuan Sen Wang, Zhi Tang, Zhi Ping Zhang, Jing Ping Liu, and Ge Lei Xiao. Hydrocephalus after aneurysmal subarachnoid hemorrhage: Epidemiology, pathogenesis, diagnosis, and management, 7 2021.

[147] Asad Ikram, Muhammad Ali Javaid, Santiago Ortega-Gutierrez, Magdy Selim, Sarah Kelangi, Syed Muhammad Hamza Anwar, Michel T. Torbey, and Afshin A. Divani. Delayed Cerebral Ischemia after Subarachnoid Hemorrhage, 11 2021.

[148] Christine Rodriguez-Régent, Monia Hafsa, Guillaume Turc, Wagih Ben Hassen, Myriam Edjlali, Alain Sermet, Nathalie Laquay, Denis Trystram, Fawaz Al-Shareef, Jean Francois Meder, Bertrand Devaux, Catherine Oppenheim, and Olivier Naggara. Early quantitative CT perfusion parameters variation for prediction of delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage. *European Radiology*, 26(9):2956–2963, 9 2016.

[149] Yasser B. Abulhasan, Najayeb Alabdulraheem, Gabrielle Simoneau, Mark R. Angle, and Jeanne Teitelbaum. Mortality after Spontaneous Subarachnoid Hemorrhage: Causality and Validation of a Prediction Model. *World Neurosurgery*, 112:e799–e811, 4 2018.

[150] Simone A. Dijkland, Bob Roozenbeek, Patrick A. Brouwer, Hester F. Lingsma, Diederik W. Dippel, Leonie J. Vergouw, Mervyn D. Vergouwen, and Mathieu Van Der Jagt. Prediction of 60-day case fatality after aneurysmal subarachnoid hemorrhage:

External validation of a prediction model. *Critical Care Medicine*, 44(8):1523–1529, 8 2016.

[151] Jordi de Winkel, Tim Y. Cras, Ruben Dammers, Pieter Jan van Doormaal, Mathieu van der Jagt, Diederik W.J. Dippel, Hester F. Lingsma, and Bob Roozenbeek. Early predictors of functional outcome in poor-grade aneurysmal subarachnoid hemorrhage: a systematic review and meta-analysis. *BMC Neurology*, 22(1), 12 2022.

[152] Masahito Katsuki, Shin Kawamura, and Akihito Koh. Easily Created Prediction Model Using Automated Artificial Intelligence Framework (Prediction One, Sony Network Communications Inc., Tokyo, Japan) for Subarachnoid Hemorrhage Outcomes Treated by Coiling and Delayed Cerebral Ischemia. *Cureus*, 6 2021.

[153] Benjamin W.Y. Lo, R. Loch Macdonald, Andrew Baker, and Mitchell A.H. Levine. Clinical outcome prediction in aneurysmal subarachnoid hemorrhage using bayesian neural networks with fuzzy logic inferences. *Computational and Mathematical Methods in Medicine*, 2013, 2013.

[154] Nora Franziska Dengler, Vince Istvan Madai, Meike Unteroberdörster, Esra Zihni, Sophie Charlotte Brune, Adam Hilbert, Michelle Livne, Stefan Wolf, Peter Vajkoczy, and Dietmar Frey. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores.

[155] Paula de Toledo, Pablo M. Rios, Agapito Ledezma, Araceli Sanchis, Jose F. Alen, and Alfonso Lagares. Predicting the outcome of patients with subarachnoid hemorrhage using machine learning techniques. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):794–801, 2009.

[156] Isabel Fragata and Patrícia Canhão. Imaging predictors of outcome in acute spontaneous subarachnoid hemorrhage: a review of the literature, 2 2019.

[157] Sergio García-García, Santiago Cepeda, Dominik Müller, Alejandra Mosteiro, Ramón Torné, Silvia Agudo, Natalia de la Torre, Ignacio Arrese, and Rosario Sarabia. Mor-

tality Prediction of Patients with Subarachnoid Hemorrhage Using a Deep Learning Model Based on an Initial Brain CT Scan. *Brain Sciences*, 14(1), 1 2024.

[158] A Hijdra, J Van Gijn, N J D Nagelkerke, M Vermeulen, and H Van Crevel. Prediction of Delayed Cerebral Ischemia, Rebleeding, and Outcome After Aneurysmal Subarachnoid Hemorrhage. Technical report.

[159] Jan Claassen, Gary L Bernardini, Kurt Kreiter, Joseph Bates, Yunling E Du, Daphne Copeland, E Sander Connolly, and Stephan A Mayer. Effect of Cisternal and Ventricular Blood on Risk of Delayed Cerebral Ischemia After Subarachnoid Hemorrhage The Fisher Scale Revisited. Technical report, 2001.

[160] Bappaditya Ray, Stephen R. Ross, Gopichand Danala, Faranak Aghaei, Claire Delpirou Nouh, Lance Ford, Kimberly M. Hollabaugh, Brittany N. Karfonta, Joshua A. Santucci, Benjamin O. Cornwell, Bradley N. Bohnstedt, Bin Zheng, George L. Dale, and Calin I. Prodan. Systemic response of coated-platelet and peripheral blood inflammatory cell indices after aneurysmal subarachnoid hemorrhage and long-term clinical outcome. *Journal of Critical Care*, 52:1–9, 8 2019.

[161] Fana Alemseged, Thanh N. Nguyen, Francisco Mont Alverne, Xinfeng Liu, Wouter J. Schonewille, and Raul G. Nogueira. Endovascular Therapy for Basilar Artery Occlusion, 4 2023.

[162] J H Warwick Pexman, Philip A Barber, Michael D Hill, Robert J Sevick, Andrew M Demchuk, Mark E Hudon, William Y Hu, and Alastair M Buchan. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT Scans in Patients with Acute Stroke. Technical report, 2001.

[163] Volker Puetz, Andrei Khomenko, Michael D. Hill, Imanuel Dzialowski, Patrik Michel, Christian Weimar, Christine A.C. Wijman, Heinrich P. Mattle, Stefan T. Engelter, Keith W. Muir, Thomas Pfefferkorn, David Tanne, Kristina Szabo, L. Jaap Kappelle, Ale Algra, Ruediger Von Kummer, Andrew M. Demchuk, and Wouter J. Schonewille.

Extent of hypoattenuation on CT angiography source images in basilar artery occlusion: Prognostic value in the Basilar Artery International Cooperation Study. *Stroke*, 42(12):3454–3459, 12 2011.

[164] Behzad Farzin, Robert Fahed, Francois Guilbert, Alexndre Y. Poppe, Nicole Daneault, André P Durocher, Sylvian Lanthier, Hayet Boujhdani, Naim N Khoury, Daniel Roy, Alain Weill, Jean-Christophe Gentric, André L Batista, Laurent L{\'e}tourneau, Francois Bergeron, Marc-Antonie Henry, Tim E. Darsaut, and Jean Raymond. Early CT changes in patients admitted for thrombectomy Intrarater and interrater agreement. Technical report, ANN Entrerprises, 2016.

[165] Alexis T. Wilson, Sadanand Dey, James W. Evans, Mohamed Najm, Wu Qiu, and Bijoy K. Menon. Minds treating brains: understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke, 2 2018.

[166] Dong Hoon Shin, Dong Jin Shin, and Jung Ryul Kim. Do All ASPECT Score Regions have the Same Predictive Power for Functional Outcomes? *Journal of Stroke and Cerebrovascular Diseases*, 29(2):104516, 2 2020.

[167] Julian Schröder and Götz Thomalla. A critical review of Alberta stroke program early CT score for evaluation of acute stroke imaging, 1 2017.

[168] Reuben Dorent, Thomas Booth, Wenqi Li, Carole H. Sudre, Sina Kafiabadi, Jorge Cardoso, Sebastien Ourselin, and Tom Vercauteren. Learning joint segmentation of tissues and brain lesions from task-specific hetero-modal domain-shifted datasets. *Medical Image Analysis*, 67, 1 2021.

[169] Hamed Asadi, Richard Dowling, Bernard Yan, and Peter Mitchell. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS ONE*, 9(2), 2 2014.

[170] Wan Yin Lin, Chun Hsien Chen, Yi Ju Tseng, Yu Ting Tsai, Ching Yu Chang, Hsin Yao Wang, and Chih Kuang Chen. Predicting post-stroke activities of daily living through

a machine learning-based approach on initiating rehabilitation. *International Journal of Medical Informatics*, 111:159–164, 3 2018.

[171] Yiran Zhou, Di Wu, Su Yan, Yan Xie, Shun Zhang, Wenzhi Lv, Yuanyuan Qin, Yufei Liu, Chengxia Liu, Jun Lu, Jia Li, Hongquan Zhu, Weiyin Vivian Liu, Huan Liu, Guiling Zhang, and Wenzhen Zhu. Feasibility of a Clinical-Radiomics Model to Predict the Outcomes of Acute Ischemic Stroke. *Korean Journal of Radiology*, 23(8):811–820, 8 2022.

[172] Albert J. Yoo, Zeshan A. Chaudhry, Raul G. Nogueira, Michael H. Lev, Pamela W. Schaefer, Lee H. Schwamm, Joshua A. Hirsch, and R. Gilberto González. Infarct volume is a pivotal biomarker after intra-arterial stroke therapy. *Stroke*, 43(5):1323–1330, 5 2012.

[173] Lauren A. Beslow, Arastoo Vossough, Hisham M. Dahmoush, Sudha Kilaru Kessler, Rebecca Stainman, Christopher G. Favilla, Courtney J. Wusthoff, Sarah Zelonis, Daniel J. Licht, Rebecca N. Ichord, and Sabrina E. Smith. Modified pediatric ASPECTS correlates with infarct volume in childhood arterial ischemic stroke. *Frontiers in Neurology*, JUL, 2012.

[174] Marina Padroni, Andrea Bernardoni, Carmine Tamborino, Gloria Roversi, Massimo Borrelli, Andrea Saletti, Alessandro De Vito, Cristiano Azzini, Luca Borgatti, Onofrio Marcello, Christopher D'Esterre, Stefano Ceruti, Ilaria Casetta, Ting Yim Lee, and Enrico Fainardi. Cerebral blood volume ASPECTS is the best predictor of clinical outcome in acute ischemic stroke: A retrospective, combined semi-quantitative and quantitative assessment. *PLoS ONE*, 11(1), 1 2016.

[175] E. François Aldrich, Randall Higashida, Abdel Hmissi, Elizabeth J. Le, R. Loch Macdonald, Angelina Marr, Stephan A. Mayer, Sébastien Roux, and Nicolas Bruder. Thick and diffuse cisternal clot independently predicts vasospasm-related morbidity and poor outcome after aneurysmal subarachnoid hemorrhage. *Journal of Neurosurgery*, 134(5):1553–1561, 5 2021.

[176] Gabriela Trifan, Baback Arshi, and Fernando D. Testai. Intraventricular Hemorrhage Severity as a Predictor of Outcome in Intracerebral Hemorrhage. *Frontiers in Neurology*, 10, 3 2019.

[177] Gopichandh Danala, Masoom Desai, Bappaditya Ray, Morteza Heidari, Sai Kiran R. Maryada, Calin I. Prodan, and Bin Zheng. Applying Quantitative Radiographic Image Markers to Predict Clinical Complications After Aneurysmal Subarachnoid Hemorrhage: A Pilot Study. *Annals of Biomedical Engineering*, 50(4):413–425, 4 2022.

[178] Jennifer A Frontera, Jan Claassen, J Michael Schmidt, Katja E Wartenberg, Richard Temes, E Sander Connolly, R Loch Macdonald, and Stephan A Mayer. Prediction of symptomatic vasospasmafter subarachnoid hemorrhage: the modified fisher scale. *Neurosurgery*, 59(1):21–27, 2006.

[179] Samir S. Yadav and Shivajirao M. Jadhav. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6(1), 12 2019.

[180] Zahangir Alom, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. Technical report.

[181] Yuliana Jiménez Gaona and Cayetano Alto. DenseNet for Breast Tumor Classification in Mammographic Images. Technical report.

[182] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional Block Attention Module. Technical report.

[183] Tamilarasi Suresh, Tsehay Admassu Assegie, Subhashni Rajkumar, and Napa Komal Kumar. A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model. *International Journal of Electrical and Computer Engineering*, 12(2):1831–1838, 4 2022.

[184] Protima Khan, Md Fazlul Kader, S. M.Riazul Islam, Aisha B. Rahman, Md Shahriar Kamal, Masbah Uddin Toha, and Kyung Sup Kwak. Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances. *IEEE Access*, 9:37622–37655, 2021.

[185] Mohammed G. Alsubaie, Suhuai Luo, and Kamran Shaukat. Alzheimer's Disease Detection Using Deep Learning on Neuroimaging: A Systematic Review. *Machine Learning and Knowledge Extraction*, 6(1):464–505, 2 2024.

[186] Ritu Gautam and Manik Sharma. Prevalence and Diagnosis of Neurological Disorders Using Different Deep Learning Techniques: A Meta-Analysis. *Journal of Medical Systems*, 44(2), 2 2020.

[187] Gopichandh Danala, Sai Kiran Reddy Maryada, Morteza Heidari, Bappaditya Ray, Masoom Desai, and Bin Zheng. A new interactive visual-aided decision-making supporting tool to predict severity of acute ischemic stroke. page 66. SPIE-Intl Soc Optical Eng, 2 2020.

[188] Gopichandh' 'Ray Bappaditya' 'Islam Warid' 'Ebert David' 'Faiz, Rowzat' 'Danala. Evaluation of deep learning framework coupled with interactive user interface to predict clinical complications after Aneurysmal Subarachnoid Hemorrhage. In *SPIE Medical Imaging 2024: Clinical and Biomedical Imaging;12930(2024)*, San Diego, 2024.

[189] H. Baradaran, V. Fodera, D. Mir, K. Kesavobhotla, J. Ivanidze, U. Ozbek, A. Gupta, J. Claassen, and P. C. Sanelli. Evaluating CT perfusion deficits in global cerebral edema after aneurysmal subarachnoid hemorrhage. *American Journal of Neuroradiology*, 36(8):1431–1435, 8 2015.

[190] Megan E. Buechel, Danielle Enserro, Robert A. Burger, Mark F. Brady, Katrina Wade, Angeles Alvarez Secord, Andrew B. Nixon, Seyedehnafiseh Mirniaharikandehei, Hong Liu, Bin Zheng, David M. O'Malley, Heidi Gray, Krishnansu S. Tewari, Robert S. Mannel, Michael J. Birrer, and Kathleen N. Moore. Correlation of imaging

and plasma based biomarkers to predict response to bevacizumab in epithelial ovarian cancer (EOC). *Gynecologic Oncology*, 161(2):382–388, 5 2021.

[191] Victor Dumas, Killian Martin, Clément Giraud, Julia Prigent, William Bloch, Karim Soualmi, Guillaume Herpe, Samy Boucebci, Jean Philippe Neau, Rémy Guillevin, and Stéphane Velasco. Functional outcome in low-ASPECTS (0–5) acute ischemic stroke treated with mechanical thrombectomy: impact of laterality explored in a single-center study. *Frontiers in Neurology*, 14, 2023.

[192] Helge C. Kniep, Sarah Elsayed, Jawed Nawabi, Gabriel Broocks, Lukas Meyer, Matthias Bechstein, Noel Van Horn, Marios Psychogios, Götz Thomalla, Fabian Flottmann, André Kemmling, Susanne Gellißen, Jens Fiehler, Peter B. Sporns, and Uta Hanning. Imaging-based outcome prediction in posterior circulation stroke. *Journal of Neurology*, 269(7):3800–3809, 7 2022.

[193] G. J. Del Zoppo, R. Von Kummer, and G. F. Hamann. Ischaemic damage of brain microvessels: Inherent risks for thrombolytic treatment in stroke, 1998.

[194] Ahmed Esmael, Mohammed Elsherief, Ahmed Abdel Khalek Abdel Razek, Nehal Tharwat Mohammed El-Sayed, Mohammed Abd Elsalam, Mohamed E. Flifel, and Saad Shawki. Relationship of Alberta Stroke Program Early CT Score (ASPECTS) with the outcome of ischemic stroke and the neurocognitive stroke biomarkers. *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 57(1), 12 2021.

[195] Tomohisa Nezu, Masatoshi Koga, Jyoji Nakagawara, Yoshiaki Shiokawa, Hiroshi Yamagami, Eisuke Furui, Kazumi Kimura, Yasuhiro Hasegawa, Yasushi Okada, Satoshi Okuda, Kazuomi Kario, Masaki Naganuma, Koichiro Maeda, Kazuo Minematsu, and Kazunori Toyoda. Early ischemic change on CT versus diffusion-weighted imaging for patients with stroke receiving intravenous recombinant tissue-type plasminogen activator therapy: Stroke acute management with urgent risk-factor assessment and improvement (SAMURAI) rt-PA registry. *Stroke*, 42(8):2196–2200, 8 2011.

[196] Hideaki Tei, Shinichiro Uchiyama, Toru Usui, and Kuniko Ohara. Posterior circulation ASPECTS on diffusion-weighted MRI can be a powerful marker for predicting functional outcom. *Journal of Neurology*, 257(5):767–773, 5 2010.

[197] Wenjun Liang, Angela T.H. Kwan, Huifang Ye, Mohammad Mofatteh, Mingzhu Feng, Jack Wellington, Pingzhong Fu, Wenlong Wei, Yu Sun, Jianhui Huang, Jie Luo, Yimin Chen, Shuiquan Yang, and Sijie Zhou. Post-ASPECTS and Post-PC-ASPECTS Predict the Outcome of Anterior and Posterior Ischemic Stroke Following Thrombectomy. *Risk Management and Healthcare Policy*, 16:2757–2769, 2023.

[198] Hiroaki Hashimoto, Tomoyuki Maruo, Yuki Kimoto, Masami Nakamura, Takahiro Fujinaga, Hajime Nakamura, and Yukitaka Ushio. The association between diffusion-weighted imaging-Alberta Stroke Program Early Computed Tomography Score and the outcome following mechanical thrombectomy of anterior circulation occlusion. *Interdisciplinary Neurosurgery: Advanced Techniques and Case Management*, 33, 9 2023.

[199] Muhib Khan, Grayson L. Baird, Richard P. Goddeau, Brian Silver, and Nils Henninger. Alberta stroke program early CT score infarct location predicts outcome following M2 occlusion. *Frontiers in Neurology*, 8(MAR), 3 2017.

[200] Richard Beare, Jian Chen, Thanh G. Phan, Kennedy R. Lees, Myzoon Ali, A. Alexandrov, P. M. Bath, E. Bluhmki, N. Bornstein, L. Claesson, S. M. Davis, G. Donnan, H. C. Diener, M. Fisher, M. Ginsberg, B. Gregson, J. Grotta, W. Hacke, M. G. Hennerici, M. Hommel, M. Kaste, P. Lyden, J. Marler, K. Muir, R. Sacco, A. Shuaib, P. Teal, N. G. Wahlgren, S. Warach, and C. Weimar. Googling Stroke ASPECTS to determine disability: Exploratory analysis from vista-acute collaboration. *PLoS ONE*, 10(5), 5 2015.

[201] Seong Joon Lee, Gyuha Park, Dohyun Kim, Sumin Jung, Soohwa Song, Ji Man Hong, Dong Hoon Shin, and Jin Soo Lee. Clinical evaluation of a deep-learning model for automatic scoring of the Alberta stroke program early CT score on non-contrast CT. *Journal of NeuroInterventional Surgery*, 16(1):61–66, 4 2023.

[202] Adelina Yafasova, Emil Loldrup Fosbøl, Søren Paaske Johnsen, Christina Kruuse, Jeppe Kofoed Petersen, Amna Alhakak, Naja Emborg Vinding, Christian Torp-Pedersen, Gunnar Hilmar Gislason, Lars Køber, and Jawad Haider Butt. Time to Thrombolysis and Long-Term Outcomes in Patients With Acute Ischemic Stroke: A Nationwide Study. *Stroke*, 52(5):1724–1732, 5 2021.

[203] R.. Rajeswari, J.. Satheesh Kumar, and T.. Devi. *LASSO: A Feature Selection Technique In Predictive Modeling For Machine Learning*. IEEE, 2016.

[204] Lei Fan, Shuai Chen, Qun Li, and Zhouli Zhu. *Variable Selection and Model Prediction Based on Lasso, Adaptive Lasso and Elastic Net*.

[205] Seyedehnafiseh Mirniaharikandehei, Morteza Heidari, Gopichandh Danala, Sivaramakrishnan Lakshmivarahan, and Bin Zheng. Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using CT images. *Computer Methods and Programs in Biomedicine*, 200, 3 2021.