UNIVERSITY OF TURKU

A Robust Deep Learning Model for Predicting Gestational Age Based on Neonates Electroencephalogram (EEG)

Health Technology
Master’s Degree Programme in Health Technology
Department of Computing, Faculty of Technology
Master of Science in Technology Thesis

Author:
Donya Abdolzadehgan

Supervisors:
Assoc. Prof. Jetro Tuulari
Assoc. Prof. Antti Airola
M.D. Silja Luotonen

June 2024

Master of Science in Technology Thesis
Department of Computing, Faculty of Technology
University of Turku

Subject: Health Technology
Programme: Master's Degree Programme in Health Technology
Author: Donya Abdolzadehgan
Title: A Robust Deep Learning Model for Predicting Gestational Age Based on Neonates Electroencephalogram (EEG)
Number of pages: 57 pages
Date: June 2024

The early stages of life play a pivotal role in shaping neurodevelopment, and understanding the factors influencing neonatal brain maturation is crucial for assessing long-term outcomes. Traditional methods, such as Ultrasound Scans (USS) and the Ballard Score, present limitations in accessibility and objectivity, especially in resource-constrained settings. This research proposes a novel approach to predicting gestational age in neonates by leveraging the time-series data of 16-channel EEG recordings and a hybrid CNN-LTM architecture. Unlike existing studies that have primarily focused on specific aspects of prenatal maternal conditions, this research shifts the focus to the direct assessment of neurodevelopment using EEG data. By predicting gestational age, the study endeavours to contribute valuable insights into the dynamic changes occurring in the neonatal brain.

Inspired by the success of deep learning in various medical imaging and time-series analysis tasks, this research seeks to harness the power of neural networks for predicting brain age. The study aims to evaluate the suggested model's efficiency by evaluating its performance against traditional CNN and LSTM models, as well as assessing its predictive capabilities in the context of existing literature on gestational age prediction. The model was trained and evaluated using a dataset of EEG recordings from neonates with chronological age 0-5 days, with performance metrics including MAE, RMSE, and R^2. Results demonstrate the model's ability to accurately predict gestational age, with strong correlations between predicted and actual values (MAE=3.16 days, RMSE=4.38 days, and R^2= 0.75). Advantages of the proposed approach include robust performance and potential utility in clinical settings, while limitations such as interpretability and generalizability are also acknowledged. Future research directions include exploring additional data modalities and addressing model limitations to further advance the field of gestational age prediction. Overall, this study contributes to the development of accurate and reliable predictive models for neonatal care.

Keywords: Gestational Age Prediction, Electroencephalogram, Deep Learning, Long Short-Term Memory, Convolutional Neural Network

Table of contents

Table of Figures

Table of Tables

1 Introduction

Precise determination of gestational age, referring to the duration from the initial day of a woman's last menstrual period to the present date, is vital for making critical clinical decisions regarding newborn care and for advancing perinatal health research. There are different common gestational age prediction methods during pregnancy such as ultrasound dating or a woman's last menstrual period [1]. While ultrasound dating offers greater accuracy compared to relying solely on a woman's last menstrual period, its widespread implementation is hindered by accessibility issues in many developing regions and among women with inadequate prenatal care. Moreover, the utility of gestational age dating through fetal ultrasound is compromised when dealing with neonates who exhibit extreme variations in size relative to their expected gestational age [2]. This underscores the need for alternative approaches to gestational age determination, especially in areas with limited resources and for populations with limited access to prenatal services. Moreover, there are some postnatal examinations such as birth weight and standardized scoring systems based on physical and neuromuscular attributes of the neonate, as also metabolite screening data, while they all have poor reliability and precision [3]. Hence, there is a pressing need to explore new and accurate approaches for estimating gestational age at birth, especially in regions lacking access to routine ultrasound dating. Such methods are essential for monitoring preterm birth rates and ensuring appropriate neonatal care in diverse healthcare settings.

Electroencephalography (EEG) emerges as a promising tool for predicting gestational age after birth, offering several benefits over traditional methods. EEG, which records electrical activity in the brain, provides a non-invasive and objective measure that can potentially reflect neurological maturity, thereby serving as a proxy for gestational age [4]. EEG recording provides a more practical and accessible option to ultrasonography dating, which depends on specialized equipment and precise memory of menstrual cycle dates [5]. Because it does away with the necessity for expensive ultrasound equipment and accurate menstrual cycle tracking, it is especially helpful in environments with limited resources. Additionally, EEG has the potential to offer insights into brain development and function, which are intricately linked to gestational age and neonatal outcomes [5]. By leveraging advanced machine learning algorithms and signal processing techniques, EEG data can be analyzed to develop accurate models for predicting gestational age at birth. Such EEG-based approaches hold promise for improving the precision and reliability of gestational age estimation, particularly in populations where traditional methods may be unreliable or inaccessible [6]. Therefore, EEG can represent a valuable avenue for advancing the field of perinatal medicine, addressing the critical need for accurate gestational age estimation in diverse healthcare settings.

EEG combined with machine learning (ML) techniques holds significant promise in predicting various neurological phenomena, including age-related changes in brain activity [4]. Electroencephalography (EEG) offers comprehensive temporal insights into dynamic brain functions and is utilized across multiple areas, including medical diagnostics, brain-computer interface (BCI) technology, and the analysis of sleep patterns [7]. However, the complexity of EEG signals, defined by poor signal-to-noise ratio, high dimensionality, and non-stationarity, poses challenges for accurate analysis [8]. To address these challenges, sophisticated preprocessing techniques are employed to remove artifacts and enhance the signal quality or use Deep Learning (DL) models.

Deep learning has emerged as a powerful tool for analyzing EEG data, offering a promising alternative to traditional ML methods. DL models, particularly deep neural networks, have the capacity to automatically learn hierarchical representations from raw EEG signals [9], enabling them to capture complex patterns and relationships that may be indicative of age-related changes in brain activity. Moreover, by leveraging large-scale EEG datasets and advanced DL architectures, researchers aim to develop robust models for accurately predicting age based on EEG signals [10]. DL models have the potential to generalize well across different populations and settings, making them valuable tools for age prediction in diverse healthcare contexts [11]. By training deep neural networks on diverse EEG datasets representing various age groups and demographic characteristics, researchers are able to create models that are robust and reliable across different populations. As a result, one of the key advantages of DL in EEG processing is its ability over conventional ML algorithms to manage the enormous complexity and non-linearity of EEG data. In addition to age prediction, DL-based approaches have shown promise in various other EEG-related tasks, including event detection, seizure prediction, sleep staging, and cognitive state classification [4]. The ability of deep neural networks to learn features directly from raw EEG signals streamlines the analysis process and allows for the discovery of subtle patterns that traditional methods may overlook [12].

In this thesis, I focus on developing DL models, specifically the combined Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture, for predicting gestational age based on neonate EEG signals. By harnessing the power of deep learning, I aim to address the challenges associated with traditional methods of gestational age estimation, particularly in resource-limited settings and for populations with limited access to prenatal services. The proposed CNN-LSTM model offers several advantages, including the ability to automatically learn hierarchical representations from raw EEG data, handle the high dimensionality and non-linearity of EEG signals, and generalize well across different populations and settings. Additionally, by leveraging large-scale EEG datasets and advanced DL architectures, the CNN-LSTM model has the potential to improve the precision and reliability of gestational age prediction, ultimately enhancing neonatal care and perinatal health outcomes.

1.1 Research Questions

The goal of this study is to look into the following research topics about gestational age prediction:

- How can electroencephalogram data be leveraged to develop accurate predictive models for estimating gestational age at birth?
- What are the advantages and limitations of using EEG data and deep learning techniques for gestational age prediction?
- Can the proposed predictive model be translated into practical applications in clinical settings to assist healthcare professionals in estimating gestational age and informing clinical decision-making processes?

Valuable patterns of brain activity can be observed from EEG data, which may provide a strong predictor of gestational age at birth. Using advanced machine learning techniques, including deep learning architectures, EEG signals can be investigated for identifying patterns indicative of fetal development stages [6]. This study will, therefore, seek to explore the interaction between EEG features and gestational age. We use deep learning to tap into the vast EEG information repository for constructing robust predictive models. We explore the predictive power underlying the EEG data through comprehensive data preprocessing and model training and evaluation in order to develop appropriate models to estimate gestational age accurately.

EEG data combined with deep learning techniques offer several advantages for gestational age prediction, including non-invasiveness, the ability to monitor in real-time, and the possibility of early detection of developmental abnormalities. Accurate gestational age prediction is made possible by deep learning models' ability to automatically extract complicated features from EEG data. However, challenges such as the interpretability of deep learning models, generalizability across different patient populations, and potential biases in the training data need to be addressed. Furthermore, deep learning algorithms' resources and computational complexity may provide real-world challenges in healthcare contexts.

The ultimate goal of this research is to develop a predictive model that can be seamlessly integrated into clinical workflows to assist healthcare professionals in estimating gestational age and making informed clinical decisions. By demonstrating the accuracy, reliability, and generalizability of the proposed model through rigorous evaluation and validation, we aim to pave the way for its practical application in real-world clinical settings. However, the translation of the model into clinical practice necessitates considerations of regulatory approval, data privacy, and user interface design tailored to the needs of healthcare professionals.

1.2 Thesis Content Summary

We undertake a thorough investigation of gestational age prediction using deep learning methods and EEG data in this thesis. The research aims to advance current methodologies and offer practical insights for clinical applications in neonatal care.

In Chapter 2, we lay the foundation for understanding the key concepts and methodologies used in this research. This chapter begins with an introduction to electroencephalography (EEG), covering its definition, applications, and the biophysics behind EEG measurements. We explore the electrode positioning system and delve into the specifics of EEG in neonates, highlighting the unique challenges and considerations in this demographic. We then discuss the methodologies for processing and analyzing EEG signals, including the identification and handling of artifacts and the overall data analysis pipeline. The chapter concludes with an overview of artificial intelligence (AI), focusing on its evolution, core concepts, and the role of machine learning and deep learning in processing large datasets and uncovering complex patterns. The relationship between AI, machine learning, and deep learning is illustrated to provide a clear conceptual framework for the thesis.

Chapter 3 provides a thorough review of existing literature related to gestational age prediction and the use of EEG data in neonatal studies. This chapter examines previous research efforts, identifying the strengths and weaknesses of various approaches. We evaluate traditional methods and modern deep learning techniques, comparing their performance and limitations. Special attention is given to studies utilizing convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, as these architectures form the basis of our proposed hybrid model. The literature review serves to contextualize our research within the broader scientific community, highlighting gaps that our study aims to fill.

The materials and methods used in our investigation are described in Chapter 4, along with participant selection, EEG data collection, preprocessing methods, and the development of the proposed CNN-LSTM hybrid model, outlining the architectural design and the rationale behind integrating CNNs and LSTMs. This chapter also discusses data balancing techniques used to address data imbalances, enhancing the model's performance and reliability. The methods section concludes with a detailed explanation of the model validation process, including cross-validation and regularization strategies to prevent overfitting and improve generalizability.

Chapter 5 presents the experimental results and a thorough discussion of our findings. First, we present the performance metrics including mean absolute error, root mean square error, and the coefficient of determination, of our proposed model on training and test datasets. These metrics are used to evaluate the accuracy and robustness of the model. We then compare our model's performance with other deep learning models and methods reported in the literature, demonstrating the advantages of our hybrid approach. The discussion section interprets the results, considering the implications for clinical practice and potential areas for improvement. We also address the limitations of our study, such as dataset biases and the need for further validation in diverse demographic settings.

In Chapter 6, we summarize the key findings of our research and their significance for predicting gestational age using EEG data. We revisit the research questions outlined in the introduction and provide detailed answers based on our experimental results. The chapter highlights the practical applications of our model in clinical settings and discusses future research directions. We emphasize the importance of further validation studies to enhance the model's generalizability and explore the integration of additional data modalities to improve prediction accuracy. The conclusion highlights the possibility of advanced deep learning methods in neonatal healthcare, aiming to contribute to better clinical outcomes and informed decision-making processes.

Finally, Chapter 7 outlines several avenues for future research to enhance the capabilities and applications of gestational age prediction models using EEG data. Potential improvements include optimizing feature selection by investigating EEG channel correlations, validating the model across diverse demographics and clinical settings, and improving model interpretability for clinicians. Incorporating preterm neonate data and developing user-friendly software for real-time predictions are also discussed to ensure practical application in neonatal care.

2 Background

2.1 Electroencephalography

2.1.1 Definition and Applications

Electroencephalography is the non-invasive neuroimaging technique mainly utilized for recording the electrical activity of the brain [13]. It can be carried out by detecting and amplifying the electrical signals from the brain with the help of electrodes on the scalp, which are displayed through waveforms on a monitor or recorded for analysis. The brain contains billions of neurons that interact with each other electrically. Whenever neurons are activated, they create electrical currents that travel through the tissue of the brain [14]. These electrical signals have the potential to develop into oscillatory patterns of activity that can be detected and recorded with EEG electrodes placed on the scalp.

The nature and characteristics of EEG signals play a crucial role in understanding brain function and diagnosing neurological disorders. Frequency, a key parameter in EEG analysis, denotes the rhythmic repetitive activity of brain waves measured in Hertz (Hz), representing the number of cycles per second [15]. Human EEG signals, originating from billions of oscillating neuron communities, exhibit a complex and unpredictable pattern characterized by intermittent bursts of oscillations [14]. The frequencies and amplitudes of these signals change in different states in healthy humans, such as wakefulness and sleep. EEG signals are typically categorized into five major brain waves based on their frequency ranges, including delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz). These frequency bands serve as markers for different brain states and cognitive processes [15].

During an EEG recording, electrodes, which are small disks, are strategically placed on different areas of the scalp using temporary adhesives. These electrodes serve as sensors to detect electrical activity in the brain. Typically, every electrode has an amplifier connected to it, one amplifier is assigned to each pair of electrodes. These amplifiers transmit the electrical signals picked up by the electrodes to an EEG recording device. The recorded signals are then processed and displayed as wavy lines on a computer screen. This graphical representation allows healthcare professionals to observe and analyze the brain's electrical activity in real time. Figure 2-1, inspired by a figure in H. Cho and J. Paik article [16], illustrates the placement of electrodes on the scalp during EEG recording and how the EEG signals appear on a computer screen. The electrodes are sensitive enough to detect minute electrical charges, in the order of microvolts (µV), generated by the brain's neurons. EEG recordings can vary in complexity, ranging from single-channel recordings to multichannel recordings with up to 256 electrodes. Each channel, typically composed of a pair of electrodes, contributes to the overall EEG signal, providing valuable insights into brain activity patterns. Subsequently, experts interpret these EEG readings to diagnose various neurological conditions and monitor brain function [15].

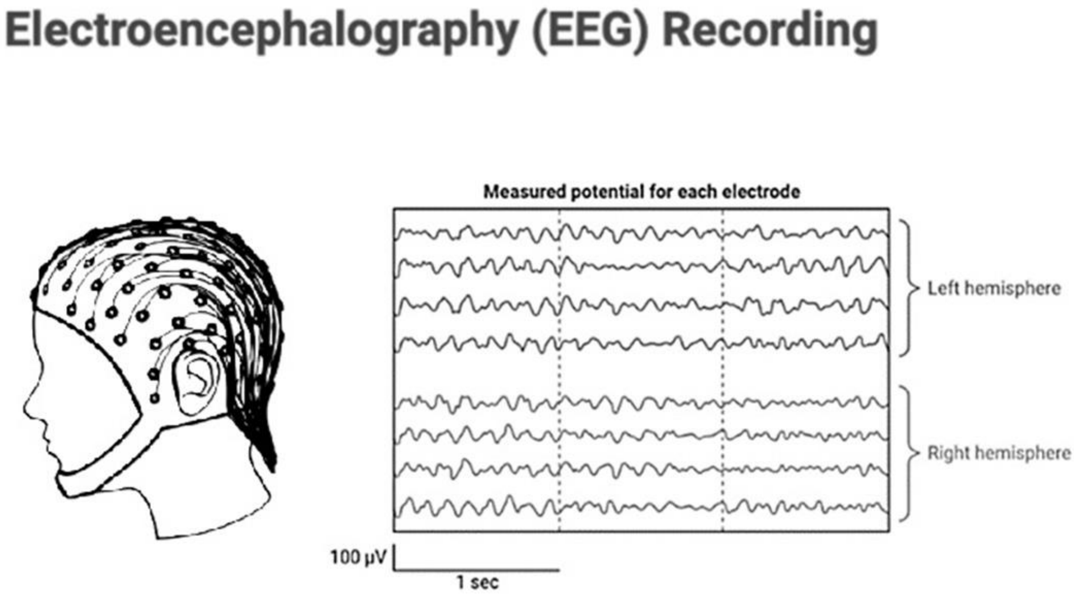

Figure 2-1. An illustration of EEG recording, Author's own drawing.

EEG finds itself applied in a wide range in the clinical, research, and engineering domains. Clinically, it is used to diagnose and monitor different neurological disorders such as epilepsy, sleep disorders, and brain injuries. In research, EEG is put to use in the evaluation of cognitive processes, emotion, and consciousness; it provides a window into the innermost workings of the human mind. In addition to that, EEG is employed in engineering applications such as brain-computer interfaces and neurofeedback systems, enabling communication between the brain and external devices [17].

2.1.2 Electroencephalography's Biophysics and Measurement

The EEG signal can be characterized by frequency, amplitude, and morphology, which usually reflect different aspects of brain functioning. The frequency of the EEG signal corresponds to the rate of neuronal oscillations, and different frequency bands are associated with certain brain states and activities. The amplitude of the EEG signal stands for the synchrony and magnitude of neuronal activity, while morphology characterizes the shape and configuration of the waveform. EEG signals are usually measured in microvolts and can be captured with the help of special EEG amplifiers and recording systems [13].

2.1.3 Electrode Positioning System

Electrode placement is crucial in EEG recording, as it determines the spatial resolution and coverage of brain activity. Electrodes are positioned according to standardized systems such as the International 10-20 system, shown in Figure 2-2, which specifies the locations for electrode placement based on anatomical landmarks on the scalp [18] [19]. The 10-20 system divides the scalp into regions based on percentages of the total distance between anatomical landmarks, ensuring consistent and reproducible electrode placement across different individuals [19].

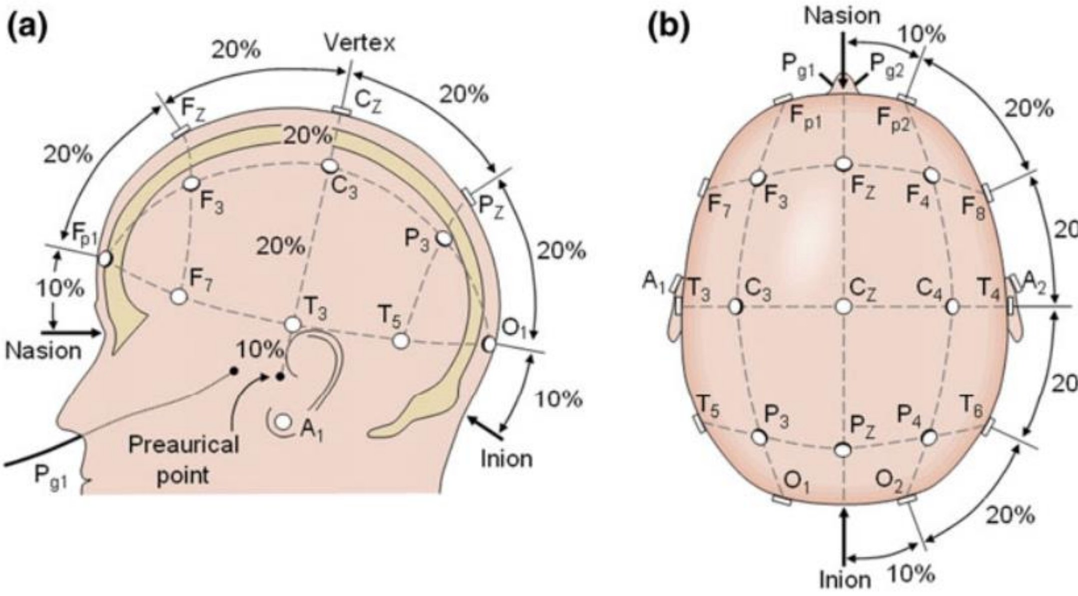

Figure 2-2. International 10–20 system EEG recording, Reprinted from The Lancet, Volume 4, K. A. Carden, Recording Sleep: The Electrodes, 10/20 Recording System, and Sleep System Specifications, Pages 333-341, Copyright (2009) [18], with permission from Elsevier.

2.2 EEG in Neonates

Neonatal EEG stands as a crucial diagnostic and prognostic tool, albeit with distinct characteristics compared to other age groups. The neonatal brain undergoes rapid maturation, resulting in EEG patterns that evolve with gestational age [20]. Serial EEG studies provide insights into expected patterns of brain maturation, aiding in the identification of deviations indicative of neurological dysfunction. Moreover, EEG findings in neonates provide useful predictive data that directs therapeutic approaches and clinical decision-making [21].

Neonatal EEG, while serving as a critical diagnostic tool, presents unique challenges and considerations distinct from EEG in adult patients. Unlike adult EEG, neonatal EEG often occurs bedside, amidst an electromagnetic noisy background, introducing numerous artifacts that complicate interpretation. Moreover, cooperation from neonatal patients is not feasible, necessitating specialized techniques for EEG acquisition and analysis [22]. Age-specific features of neonatal EEG, including specific frequencies and maturation patterns, change quickly with gestational age. Thus, effective interpretation requires an accurate understanding of gestational age [23]. The progressive decrease in the total amplitude and component power of delta activity during bursts in quiet and active sleep stages is one of the defining features of neonatal EEG maturation [24]. The emergence and maturation of sleep states play a pivotal role in neonatal brain development, reflecting processes such as thalamocortical and intracortical innervation, synaptogenesis, and synaptic remodeling. Additionally, neonatal EEG offers valuable insights into the early establishment of regulatory mechanisms governing sleep-wake states, providing further understanding of neonatal brain function and development [24].

The maturation of EEG throughout the lifespan is a dynamic process characterized by distinct developmental stages. EEG undergoes significant changes from infancy to adulthood, culminating in a stabilized pattern by approximately 30 years of age. During infancy and adolescence, there is a notable shift in EEG rhythm towards higher frequencies, reflecting ongoing brain maturation [25]. In newborns, delta rhythms predominate, gradually transitioning to theta rhythms by 12 months of age [26]. Notably, the identification of alpha rhythm around 10 years ago marked a significant milestone in EEG development. Throughout young adulthood, subtle signs of immaturity may persist in EEG patterns, including the presence of specific frequency waves during wakefulness. Physiologically, EEG maturation is intricately linked with the development of dendritic trees and myelination. Myelination, facilitated by glial cells, plays a pivotal role in insulating neuronal axons and enhancing the efficiency of electrical signal propagation [27]. Thus, understanding the maturation process of EEG provides crucial insights into brain development and functioning across different life stages.

2.3 Methodology for Processing and Analysing EEG Signals

2.3.1 EEG and Artifacts

EEG recordings, while invaluable for understanding brain activity, are often plagued by artifacts—undesirable electrical signals originating from non-cerebral sources. These artifacts can significantly affect the interpretation of EEG data, as their amplitudes may rival those of cortical signals of interest. Consequently, interpreting EEGs accurately requires considerable expertise. Common artifacts include electrooculographic artifacts induced by eye movements, electrode artifacts resulting from poor electrode-skin contact, swallowing artifacts, and artifacts from reference electrodes. These artifacts manifest as distinct waveforms on EEG recordings, such as slow positive waves in frontal electrodes or large waves across all channels. Detecting and mitigating artifacts is essential to ensure the fidelity of EEG interpretations. Various types of artifacts, including electro-galvanic signals, movement artifacts, and frequency artifacts, can distort EEG signals, necessitating their identification and removal[28]. Researchers have explored automated systems for artifact detection in EEG recordings, employing parametric methods to extract relevant features and compare them against predefined thresholds to identify and remove artifacts efficiently. Such endeavors aim to enhance the reliability and accuracy of EEG data analysis in clinical and research settings.

2.3.2 EEG Data Analysis

EEG data analysis encompasses various techniques to decipher the intricate patterns of brain activity captured by EEG recordings. Researchers employ many signal-processing methods to extract meaningful insights from EEG data, including filtering, artifact removal, feature extraction, and classification or prediction algorithms. In the time domain analysis, researchers observe the temporal characteristics of EEG waveforms, focusing on peak values and frequencies to decode brain activity or diagnose neurological conditions. Frequency domain analysis involves transforming raw EEG data into spectrum diagrams using techniques like Fourier transform or wavelet transform, enabling the identification of frequency-specific patterns indicative of different brain states. Moreover, time-frequency analysis methods combine both time and frequency domains to capture dynamic changes in EEG signals over time. Non-linear methods delve deeper into the complex interactions within EEG data, exploring patterns beyond linear relationships [28].

In recent years, deep learning techniques, particularly CNNs, have gained traction in EEG data analysis. These methods automatically extract features from raw EEG data, bypassing the need for manual feature extraction, and effectively mitigating the effects of noise, enhancing the accuracy of brain decoding and disease diagnosis. The versatility and efficacy of deep learning approaches have propelled their widespread adoption in EEG-based research and clinical applications.

2.4 Artificial Intelligence

Artificial intelligence (AI) represents a convergence of computer science and cognitive psychology, aiming to imbue machines with human-like capabilities. At its core, intelligence entails the ability to achieve goals, encompassing processes such as reasoning, problem-solving, pattern recognition, and learning from experience [29]. Within the field of artificial intelligence, machine learning focuses on creating algorithms that enable computers to learn and become more efficient over time without the need for explicit programming. [30]. ML algorithms use large-scale training data sets to discover intricate patterns and hidden insights without the need for human interaction. ML algorithms fall into three general categories: reinforcement learning, unsupervised learning, and supervised learning. Each type of algorithm is best suited for a particular set of tasks and data.

In this work, we focus on the application of supervised machine learning methods, where the model learns from labeled training data to make predictions. Specifically, we employ supervised learning techniques to develop predictive models for estimating gestational age at birth using EEG data. These models are trained on EEG data paired with corresponding gestational age labels, allowing them to learn the relationship between EEG features and gestational age through supervised learning algorithms.

Despite their effectiveness, traditional ML algorithms have limitations, particularly in handling large and high-dimensional data [31]. Deep Learning, a subfield of ML, addresses these limitations by leveraging deep neural networks (DNNs) with multiple hidden layers. Figure 2-3 shows the connection between deep learning, machine learning, and artificial intelligence. [32].

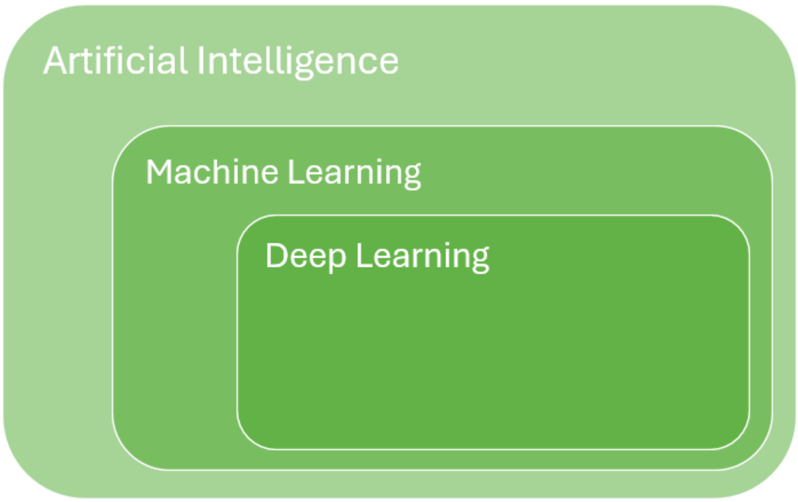

Figure 2-3. The relationship between artificial intelligence, machine learning, and deep learning, Author's own drawing.

2.4.1 Deep Learning

Deep Learning represents a paradigm shift in the realm of artificial intelligence and machine learning, transforming how computers process and interpret data. In contrast with conventional machine learning techniques, which frequently call for manual feature extraction and engineering, DL models, particularly DNNs, can automatically identify complex patterns directly from raw data [31]. This capability enables DNNs to extract intricate patterns and relationships from complex data types such as text, images, and EEG signals. To get the best results, however, deep learning models need a lot of computational resources and training data. [33].

DL models utilize advanced neurons and deeply nested architectures, enabling them to capture subtle nuances and correlations within the data. By iteratively learning from vast amounts of training data, DNNs can uncover hidden insights and complex patterns without explicit human intervention. This inherent adaptability and flexibility make DL models well-suited for several uses, including financial forecasting, medical diagnosis, computer vision, and natural language processing [33].

In the field of EEG analysis, DL has emerged as a powerful tool for processing and interpreting brain signals. Deep neural networks, such as recurrent neural networks (RNNs) and convolutional neural networks, have demonstrated remarkable efficacy in tasks such as emotion detection, seizure prediction, and cognitive state monitoring. CNNs are very effective at identifying spatial patterns in EEG data, making them particularly adept at tasks requiring spatial analysis. However, RNNs are well-suited for modeling temporal dependencies in sequential EEG data, allowing them to capture dynamic patterns and trends over time [33].

The continued advancement of DL techniques holds immense potential for the field of EEG analysis, offering new avenues for understanding brain function and developing innovative diagnostic and therapeutic strategies.

3 Literature Review

This section provides an overview of several related research articles to interpret the relationship between age and EEG signal processing.

Numerous researchers have attempted to estimate brain age from EEG signals. Some of them have shown that EEG characteristics such as rhythmic activity, vary with advancing age [34], while others have focused on feature extraction and conventional machine learning techniques. Developing an automated EEG Maturational Age (EMA) estimation for early preterm neonates is the goal of the study conducted by O'Toole et al. [35]. The EMA estimator, employing a linear combination of features, significantly outperforms a nominal reading, offering an accurate and accessible means for assessing functional brain maturity in early preterm neonates [35]. Another study introduces an automated EMA estimator based on 23 computational features, achieving a high correlation (0.936) with clinically determined postmenstrual age. All neonates showed increasing EMA with postmenstrual age, establishing EMA as a reliable surrogate measure for accurately determining brain maturation [36]. Dimitriadis & Salis (2017) utilized a feature selection technique, an Extreme Learning Machine (ELM) classifier, and an Support Vector Regressor (SVM) to distinguish between young adults and middle-aged individuals based on EEG-derived spatio-temporal features. To predict age and discriminate between groups, the study highlighted the significance of dynamically reconfiguring dominating coupling modes from EEG sensor pairs. The classification accuracy rates were notably high, with 97.8% for eyes-open and 87.2% for eyes-closed conditions when differentiating between young and middle-aged groups [37]. Sun et al. extracted 102 features from six EEG channels to create an interpretable machine learning model to predict brain age based on sleep EEG data from two large data sets. In healthy subjects, the model produces a mean absolute divergence between chronological age and brain age of 7.6 years; validation shows that brain age increases on average over time [38].

Recent advancements in machine learning techniques have opened new avenues for exploring age prediction and assessing brain age using EEG signals, presenting a less conventional yet promising alternative to MRI-based methods. Al Zoubi et al. [34] investigated the potential of leveraging EEG data in combination with ML frameworks to predict chronological age and Brain Age Gap Estimate across a cohort of 468 participants with diverse disorders. Their study showcased the efficacy of a stack-ensemble age prediction model, achieving a mean absolute error of 6.87 years, thus highlighting the viability of EEG-based age estimation methodologies [34].

The study conducted by Sun et al. [38] delves into the intricate relationship between EEG patterns during sleep and aging, proposing the concept of brain age as a measure of neurological aging. Utilizing extensive sleep EEG datasets from the Sleep Heart Health Study and Massachusetts General Hospital, the researchers created a machine learning model for predicting brain age and its divergence from chronological age. Their model demonstrated its promise as a tool for evaluating brain aging, with a mean absolute variation of 7.6 years between brain age and chronological age in healthy persons. Furthermore, a considerable rise in brain age was shown over time by longitudinal analysis, especially in individuals with diabetes, hypertension, or neurological or psychiatric problems. These findings support the use of EEG-derived brain age as a biomarker for normal aging of the brain. These results highlight the potential of using sleep EEG data to monitor and better understand age-related changes in brain function [38].

Klymenko's study explores a novel approach for classifying EEG recordings automatically, aiming to address the methodological challenges of interpreting EEG data while ensuring robustness to artifacts and varying recording durations. By adapting a method from natural language processing (NLP) and applying byte-pair encoding (BPE) to symbolize EEG signals, the study analyzed a large sample of routine clinical EEG data spanning diverse age ranges. The study used Random Forest to predict the biological age of the patients using the reconstructed EEG data. The mean absolute error of the anticipated age was 15.9 years, and the correlation between the predicted and real ages was 0.56. Additionally, the study found substantial correlations between age and the frequency of EEG patterns, especially in the frontal and occipital EEG channels. These findings highlight the potential of NLP-based approaches in EEG classification, offering insights into age prediction and facilitating the interpretation of clinical EEG data with minimal preprocessing [39].

While traditional approaches rely on predefined assumptions for feature extraction, deep learning models offer a more exploratory approach by capturing nuanced features that might be overlooked by conventional methods [40]. Stevenson et al. [41] tackled the issue of site-dependent variations in EEG-based age prediction in preterm neonates, emphasizing the critical importance of managing site differences in EEG classifiers. The research utilized a new 'bag of features' methodology, which combined feature selection and Support Vector

Regression (SVR) for predicting Post-Menstrual Age based on EEG data. By training SVR on a dataset from one site and validating it on another, they identified significant challenges in maintaining prediction accuracy across different sites due to unexpected variations in EEG signals. Notably, their findings underscored the necessity of mitigating site-dependent differences through careful feature selection and cross-validation strategies. Through their innovative approach, they shed light on the complexities of EEG interpretation in preterm neonates and proposed strategies to enhance the generalizability of EEG classifiers, paving the way toward more universal methods of age prediction in clinical practice [41].

The capacity of DL models in recent years stems from their ability to learn directly from extensive datasets, enabling automatic feature extraction [6]. In order to predict brain age, we look into different types of DL models. There are various studies suggesting employing DL models to enhance the effectiveness and precision of analyzing EEG recordings by automatically extracting crucial clinical data features. Yook et al. [42] developed a novel sleep EEG-based brain age prediction model, demonstrating superior accuracy compared to previous models. Utilizing six-channel EEG data collected over a six-hour sleep period, the researchers transformed the EEG data into 2D scalograms, which were then fed into DenseNet for brain age prediction. The model exhibited a strong correlation of 80%, with a mean absolute error of 5.4 years, between predicted brain age and chronological age. This research holds clinical relevance as the brain age index derived from the model could potentially aid in the diagnosis of individuals with sleep disorders, offering a comprehensive single index reflecting the association of different sleep disorders [42]. In another research, Kaushik et al. [43] utilized a new application of Brain-Computer Interface (BCI) technology, focusing on predicting individuals' age and gender using EEG analysis. Employing a deep BLSTM-LSTM network, the study develops a hybrid learning approach achieving notable accuracy levels of 93.7% for age prediction and 97.5% for gender identification, surpassing current methodologies. It emphasizes the effectiveness of beta band frequencies in EEG signals for this task and explores potential uses in various domains such as biometrics, healthcare, entertainment, and targeted marketing [43]. Jusseaume et al. [5] study employed deep learning techniques, specifically LSTM neural networks, to analyze EEG recordings. The proposed model, which directly utilizes raw brain wave data, demonstrates a classification accuracy of 90% for determining patients' brain ages across six distinct age groups, and achieves a mean absolute error of 7 years in age regression analysis. Importantly, the results of Bidirectional Long Short-Term Memory (BLSTM) indicate superior performance compared to the previously covered CNN models [5].

Recent advancements in deep learning methodologies have opened the door to new approaches to age prediction and brain age estimation, particularly in the realm of EEG analysis. Khayretdinova et al. [44] delved into this domain by leveraging the TD-BRAIN dataset comprising healthy controls and individuals with psychiatric disorders, a Deep Convolutional Neural Network (DCNN) model was developed. By employing data augmentation techniques and cross-validation, the model achieves a mean absolute error of 5.96 years in age prediction [44]. Another study, Ansari et al. [45], aiming to estimate neonates' biological brain age utilizes a deep learning model trained on resting-state EEG data, achieving high accuracy with a mean absolute error of 1.03 weeks and 0.98 weeks in two independent datasets. The model significantly differentiates between brain age gaps in neonates with normal and severe abnormal outcomes [45]. Ansari et al. [46] in the another investigation proposing a deep-learning approach for predicting brain age in preterm neonates using EEG, a CNN block based on the Inception architecture (Sinc) is introduced. Operating directly on EEG data, the model achieves an MAE of 0.78 weeks, distinguishing between neonates with normal and severely abnormal outcomes. [46].

Given the plethora of studies exploring the relationship between age and EEG signal processing, it becomes evident that traditional approaches have limitations, often relying on specific assumptions for feature extraction. In contrast, deep learning models offer a promising avenue for overcoming these limitations by automatically capturing elusive features from extensive datasets. Recent studies demonstrate the superiority of DL methods over traditional approaches in EEG-based age prediction. For instance, Yook et al. [42] reported a MAE of 5.4 years using a DenseNet model, significantly lower than errors typically seen in traditional ML methods. Similarly, Ansari et al. [45] achieved an MAE of 1.03 weeks using a deep learning model on neonatal EEG data, outperforming conventional ML techniques, also in another research their model distinguishes between newborns with normal and significantly aberrant outcomes, achieving an MAE of 0.78 weeks [46]. These results underscore the enhanced predictive accuracy of DL methods, which are adept at capturing complex, nonlinear relationships in EEG data that traditional methods might miss.

Furthermore, DL models like CNNs and RNNs have shown exceptional performance in feature extraction and pattern recognition without extensive preprocessing. For example, Sun et al. [38] utilized sleep EEG data to predict brain age, demonstrating an MAE of 7.6 years, and highlighted the ability of DL models to identify subtle changes in brain activity associated with aging. In another instance, Jusseaume et al. [5] achieved a high accuracy of 90% in classifying patients' brain ages using LSTM networks, indicating the robust capabilities of DL in handling temporal data.

DL approaches not only enhance accuracy but also improve generalization across diverse populations and settings. Stevenson et al. [41] addressed site-dependent variations in EEG data, emphasizing the importance of feature selection and cross-validation in maintaining prediction accuracy across different sites. This adaptability makes DL methods particularly valuable in clinical practice, where variability in data collection conditions is common.

In our study, we address these challenges by adopting a deep learning approach that circumvents the need for manual feature extraction. Specifically, we employ a combination of Convolutional Neural Network and Long Short-Term Memory architectures to predict neonate brain age based on EEG signals. This unique architecture leverages the strengths of both spatial and temporal feature extraction, thereby mitigating the reliance on complex preprocessing techniques and reducing data requirements compared to previous studies. Furthermore, we introduce data balancing techniques that are simpler and less time-consuming than typical data augmentation methods. This approach effectively addresses the data limitation challenges often faced by DL models. Through extensive experimentation, we compare the performance of CNN and LSTM models independently, ultimately determining that the fusion of these two methods yields superior results. By establishing a fully trained model, we aim to contribute to the advancement of accurate and accessible gestational age estimation in neonates, especially in settings with limited resources where conventional approaches may be impractical or unreliable.

4 Materials and Methods

The methodology comprises multiple stages, such as data preprocessing, model architecture development, training, and evaluation. We preprocess the EEG data to remove artifacts and standardize the input format. Next, we design the CNN-LSTM architecture, specifying the number of layers, neurons, and activation functions. The model is then trained on the preprocessed EEG data using appropriate optimization techniques and loss functions. Finally, we evaluate the performance of the trained model on a separate validation dataset. Through this methodology, we aim to develop a reliable and clinically applicable tool for estimating gestational age at birth using EEG signals. A summary of the framework is shown in Figure 4-1.

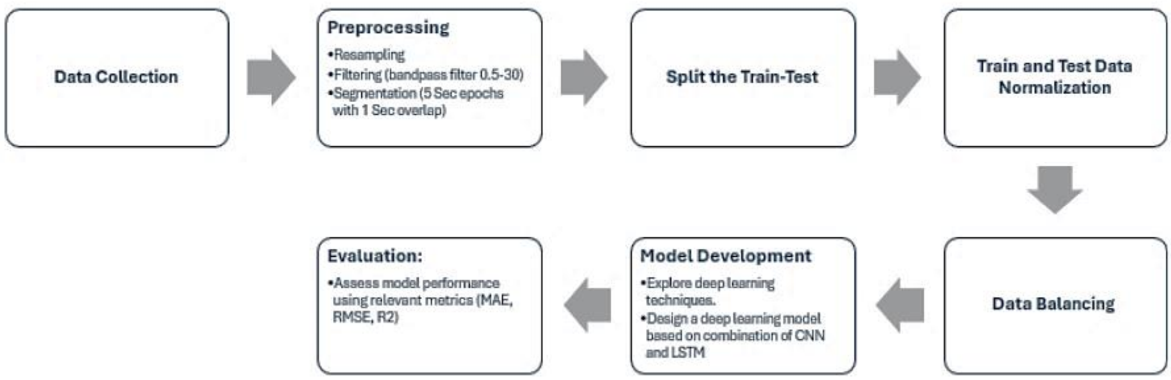

Figure 4-1. Overview of the methodology.

4.1 Participants

Individuals were selected from the FinnBrain Birth Cohort, headquartered in South-Western Finland. The aim of this cohort is to investigate the potential impacts of early life environment and genetics on neonate neurodevelopment and health [47]. During the first trimester ultrasound visit at gestational week 12, families were enlisted between December 2011 and April 2015. The FinnBrain study involved 3837 children (29 twin pairs included) in all. Adhering to the principles outlined in the Declaration of Helsinki, the EEG sub-study was approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa as well as the Ethics Committee of Southwest Finland. Prior to participation, parents gave written informed consent for their neonates to be included in the study.

In the current study, 158 neonates born between 2013 and 2015 made up the sample, recorded at Turku University Hospital's maternity wards. EEG recordings were made between one and two days after delivery. Births before gestational week 36 + 0; neonatal birth weights under 1800 grams; the existence of any severe deformity or developmental disorder; known hearing defects (which are routinely screened for all neonates); or the necessity of neonatal medical follow-up because of the mother's severe illness were among the exclusion criteria for EEG recordings. Mothers had to meet the additional requirements for inclusion, which included being proficient in Finnish or Swedish enough to accurately fill out survey questionnaires. Finally, after considering all these factors and undergoing expert scrutiny based on the visual data quality check and also data preprocessing, which included removing the first 10 seconds of recordings to prevent human artifacts at the beginning of recording and retaining recordings longer than 120 seconds, the number of neonates included in the sleep data was reduced to 76 including 36 boys and 40 girls, illustrated in Figure 4-2. Additionally, Table 4-1 provides general demographic information about the participants.

Table 4-1. Demographic information of dataset (76 Neonates)

	N	Mean	Standard Deviation	Range
Sex				
Boys	36	-	-	-
Girls	40			
Gestational Age (days)	-	280.23	8.85	253-296
Chronological Age (days)	-	1.48	1.12	0-5

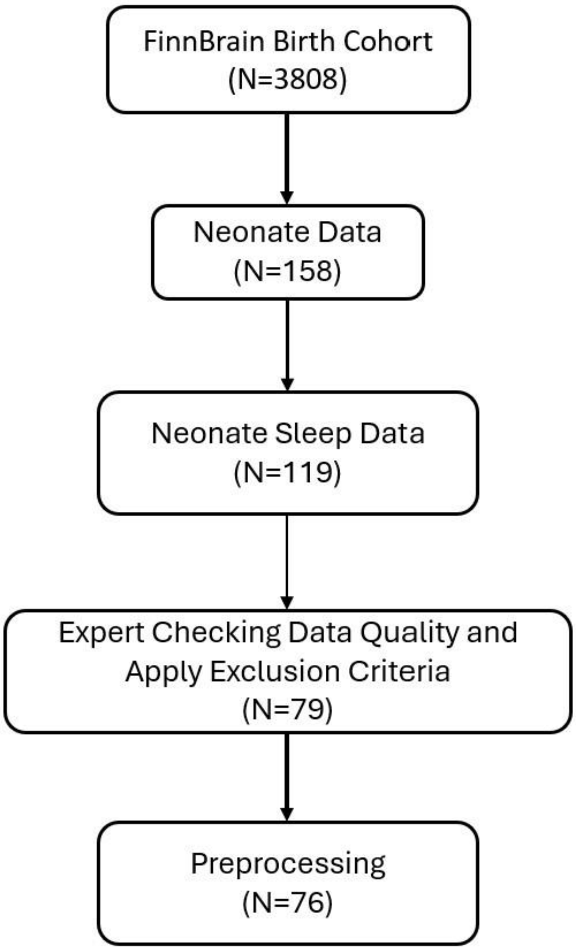

Figure 4-2. Participants Description.

4.2 EEG Data

4.2.1 EEG Recording

A sleep EEG was taken 0-5 days after giving birth. The EEG measurement was done two hours after the neonate was fed in order to minimize restlessness during sleep and to attain a tranquil sleep phase. Of the neonates, 24 percent were recorded in the afternoon, and the remaining 24 percent in the forenoon. For 16 newborns, a parent was present when the data were being collected. If the neonate remained awake after donning an EEG-headpiece (ActiCap electrode cap by EASYCAP, Germany) and applying electrode gel (Signa Gel, Parker Laboratories, Inc., USA), a dummy or a mild glucose solution, if necessary, was given to help put them to sleep. With their right ear looking upward, the newborns were sleeping on their left side. Using a BrainVision Quickamp amplifier (BrainProducts, Germany), an EEG recording with 16 channels, including Fp1, Fp2, F7, F3, Fz, F4, F8, C3, Cz, C4, P7, P3, Pz, P4, P8, and Oz (International 10–20 system) shown in Figure 4-3, was made. On the forehead were the reference and ground electrodes. The frequency of sampling was 500 Hz. A researcher assessed the neonate's level of alertness visually. During the recording, 4 (3%) and 17 (12%) of the total newborns were partially awake and in a peaceful state. The measurement session lasted for around forty-five minutes.

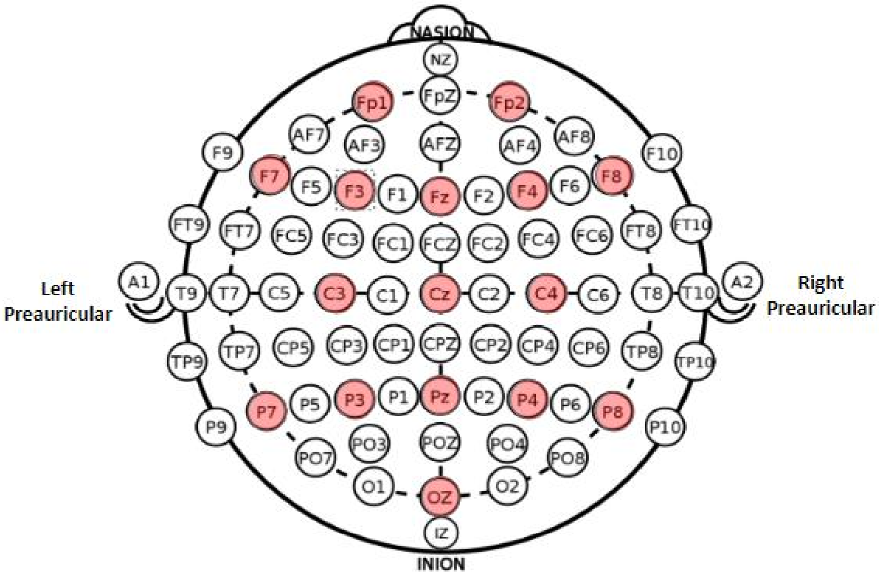

Figure 4-3. Position of the electrodes according to the 10-20 international system, Author's own drawing.

4.2.2 EEG Preprocessing

The data preprocessing process commenced with the identification and exclusion of subjects with duration of sleep recording less than 120 seconds, resulting in the exclusion of certain subjects and reducing the total number of included subjects to 76 neonates (N=76, comprising 36 boys and 40 girls). The next step was removing the first 10 seconds to avoid any potential human error. Following this initial step, each EEG recording underwent down-sampling from 500 Hz to a sampling rate of 250 Hz to standardize the temporal resolution across all recordings. Subsequently, a band-pass filter was applied to the data within the frequency range of 0.5 Hz to 30 Hz, effectively removing both low-frequency drifts and high-frequency noise artifacts. This filtering step aimed to enhance the signal quality and isolate relevant brain activity patterns from unwanted interference. Following filtering, the recordings were segmented into consecutive 5-second epochs, with a 1-second overlap between adjacent segments. This segmentation process facilitated the analysis of discrete temporal windows within the EEG data, allowing for the examination of dynamic brain activity patterns over time while minimizing data redundancy. Each segmented epoch was then subjected to subsequent analysis steps to extract features relevant to the prediction of gestational age. Figure 4-4 illustrates the difference between the raw data and pre-processed data.

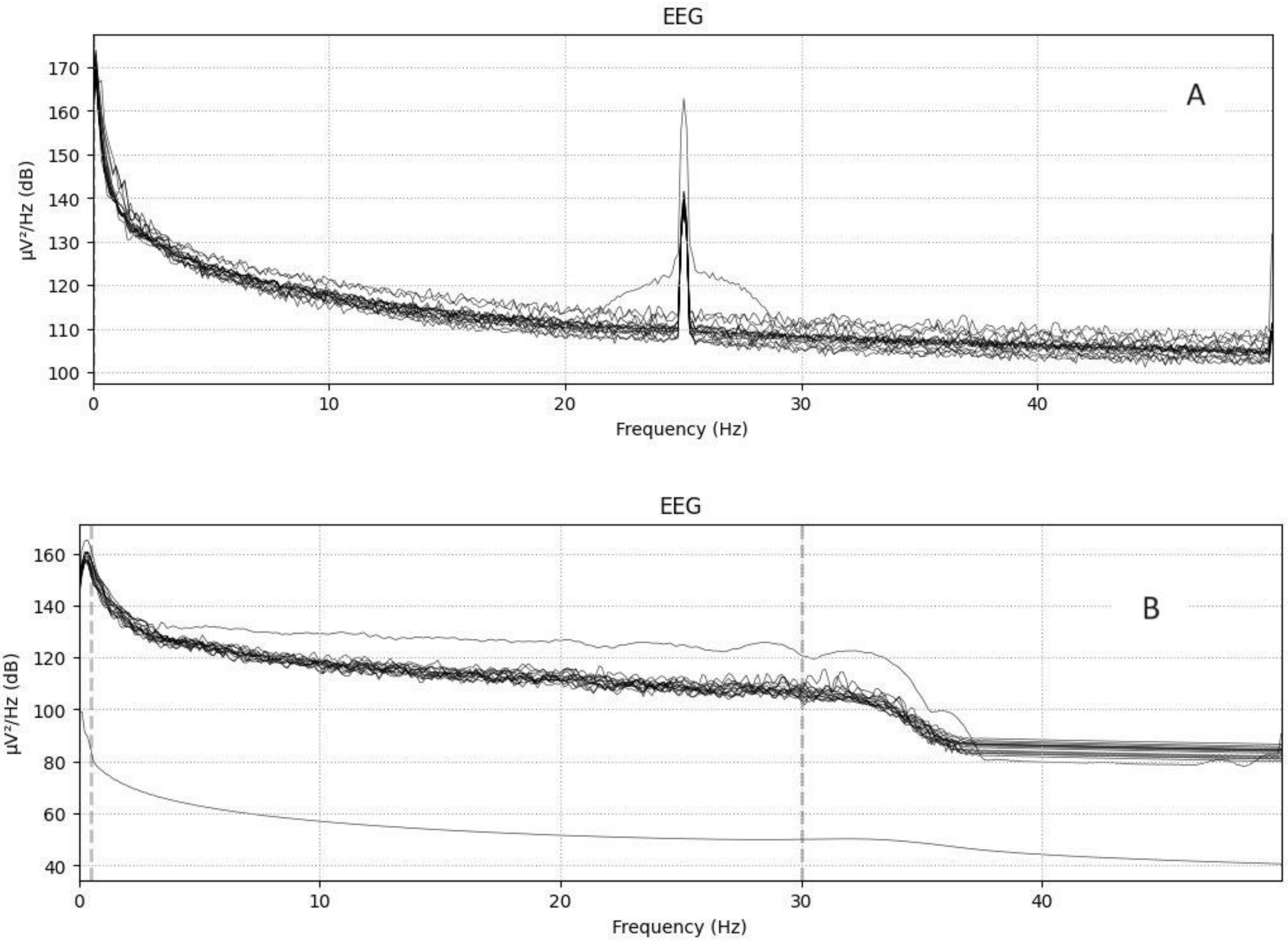

Figure 4-4. (A) Raw EEG data visualization - (B)Visualization of the pre-processed EEG data

4.3 Proposed Method

Our proposed method integrates deep learning models to solve prediction and regression problem complexities. The specific goal of our study is to use a deep neural network architecture that is capable of learning hierarchical representations automatically from raw EEG data. The architectural design basically connected multiple layers of neurons in a deep structure to enable the model to capture the intricate patterns and nonlinear relationships that underlie the EEG signals.

In order to improve the model's performance and facilitate training convergence, we incorporate preprocessing techniques such as z-score normalization. This step standardizes the features of the EEG dataset, ensuring uniformity in scale across different input variables.

In our proposed approach, the trained deep learning model will be tasked with predicting a target variable, such as gestational age, based on the input EEG data. We anticipate that deep learning models will excel in this task, particularly when dealing with complex and nonlinear correlations between EEG features and gestational age.

Moreover data balancing strategies like oversampling and undersampling were used to address any possible class imbalances in the dataset [48]. By guaranteeing a balanced distribution of classes during training, these methods avoid biases towards the majority class and enhance the model's performance in all classes.

In the domain of age prediction, we propose a specific CNN and LSTM architecture tailored to the characteristics of EEG data. This architecture is designed to capture the complex and nonlinear relationships inherent in EEG signals, utilising the advantages of recurrent and convolutional neural networks.

Cross-validation has been a fundamental technique in machine learning for assessing the robustness and generalization capability of prediction models for decades. Recent advancements in deep learning methodologies have reinforced the importance of this technique. In our study, we employed a cross-validation strategy, specifically 5-fold cross-validation, to evaluate the performance of our proposed CNN-LSTM architecture for age prediction.

The dataset was partitioned into five folds, ensuring that each fold represented a diverse distribution of samples. We then conducted cross-validation by iteratively training the model on four folds (the training set) and evaluating its performance on the remaining fold (the validation set). This process was repeated five times, with each fold serving as the validation set once, to ensure comprehensive assessment and validation of the model's performance. By integrating cross-validation, we aimed to address two key objectives. Firstly, it allowed us to assess the generalization capability of our trained model across different data partitions. Secondly, it helped us mitigate the risk of overfitting to a specific training set by evaluating the model's performance on multiple subsets of the data. It's important to note that we also reserved a separate test set, distinct from the training and validation sets, to evaluate the final model's performance independently. This test set was not used during the model training or hyperparameter tuning process, ensuring an unbiased assessment of the model's effectiveness in predicting age from EEG signals.

The best-performing model was chosen using evaluation measures including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) after a grid search for different model configurations concluded. The model with the lowest MAE, RMSE, and highest R2 values was chosen as the optimal configuration for age prediction using EEG signals. Furthermore, sensitivity analysis was performed to evaluate how robust the chosen model was to changes in the input features and hyperparameters. This involved systematically varying the model's parameters and observing changes in performance metrics, ensuring that the model's predictions remain stable and reliable across different settings.

Overall, the proposed deep learning approach, coupled with cross-validation and thorough model assessment, offers a robust and reliable framework for predicting gestational age at birth using EEG signals. By leveraging the strengths of deep learning techniques and rigorous evaluation methods, we aim to contribute to the advancement of predictive modeling in neonatal healthcare and facilitate early detection and intervention for optimal neonatal outcomes.

4.3.1 Z-score Normalization

Z-score normalization, commonly referred to as standardization, is a method used to adjust dataset features such that their mean is 0 and their standard deviation is 1. This is accomplished by deducting the mean of each feature from the respective data points and subsequently dividing by the standard deviation. The mathematical expression for z-score normalization is as follows:

$$Z = \frac{x - \mu}{\sigma}$$

Where Z is the standardized value, x is the original data point, μ is the mean of the feature, and σ is the standard deviation of the feature.

Z-score normalization is beneficial for machine learning algorithms because it ensures that all features contribute equally to the model's learning process, regardless of their original scale. It also helps in stabilizing the training process and improving convergence.

4.3.2 Data Balancing

Data balancing, an essential preprocessing step in machine learning, aims to address label imbalance within a dataset, ensuring that the model learns effectively across all target values [49]. While traditional data balancing methods are often associated with classification tasks, they can also be beneficial in regression problems like ours, where the outcome variable is continuous. In our study, the concept of a "minority label" or a "majority label" doesn't directly apply as in classification. Instead, we focused on ensuring an even distribution of target variable ranges to prevent the model from being biased towards certain age ranges.

Oversampling involves augmenting the dataset by replicating or slightly modifying instances of less frequent target values. In our context, this corresponds to increasing the representation of target value ranges with fewer instances. We accomplished this by synthetically generating additional data points for these ranges. Specifically, for target values with fewer instances, we replicated existing instances and introduced small random noise to diversify the dataset. This approach ensures that the model receives sufficient exposure to all ranges of the target variable, preventing bias towards the majority values.

Undersampling, on the other hand, aims to reduce the number of instances for target values that are excessively represented in the dataset. This prevents the model from becoming biased towards these values. To implement undersampling, we randomly selected a subset of instances from the over-represented target values. By reducing the instances of these values, we balanced the distribution of the target variable, allowing the model to learn from all categories equally.

In our methodology, we utilized a combination of oversampling and undersampling techniques to achieve a balanced dataset. Specifically, we ensured that each target value range had a comparable number of instances, thus avoiding the pitfalls of label imbalance. This approach enhances the model's performance, particularly in prediction tasks where the learning process is significantly influenced by label distribution. It's important to note that our data balancing approach was applied solely to achieve a balanced dataset and enhance model performance. We did not use it for hyperparameter optimization or any other aspect of model tuning. Instead, it served as a preprocessing step to ensure the effectiveness and reliability of our CNN-LSTM architecture for age prediction.

4.3.3 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) represent an efficient category of deep learning models that are specifically engineered to handle multi-dimensional data arrays, including time-series data, images, and audio [50]. Unlike traditional machine learning algorithms, CNNs do not necessitate the manual definition of features, as they leverage convolution kernels to automatically identify relevant local patterns directly from raw data, thereby preserving valuable information. CNNs consist of four key components: convolutional layers, pooling layers, fully connected layers, and activation functions [51]. The overall structure of CNN is shown in Figure 4-5 [52].

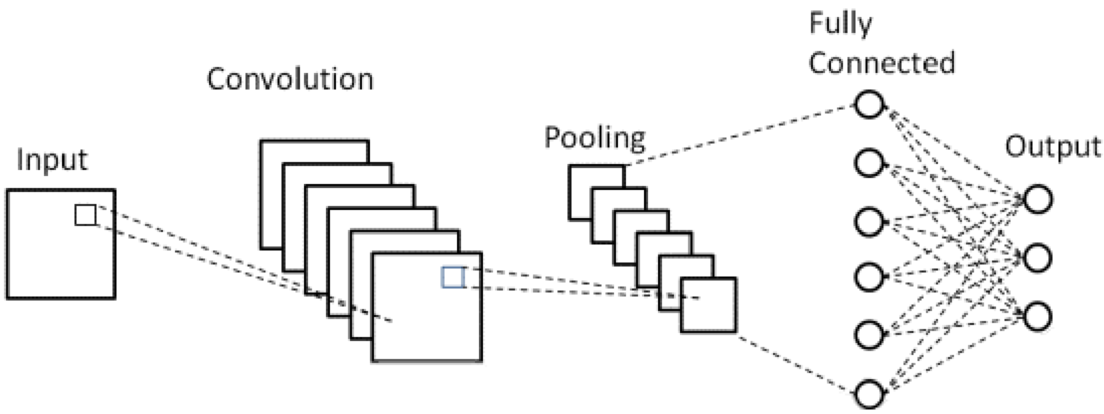

Figure 4-5. The CNN architecture, Reprinted from "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets" by Phung and Rhee. (2019) [52], which is licensed under a Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).

The fundamental ideas of CNN design are shared weights and local connections, which are especially useful for handling time series data [4]. CNNs are useful in identifying local patterns or motifs that point to age-related changes in brain activity within the context of EEG data, which may be seen as a series of data points across time.

A standard CNN design for time series signals consists of stacking layers for pooling, non-linearity, and convolution at various stages [53]. With the help of this hierarchical structure, the network is able to identify both local and global patterns in the EEG signal and learn progressively abstract representations of the input data. CNNs are able to update each weight in the filter banks by backpropagating gradients through the network, which enables the model to adjust to the underlying patterns found in the EEG data [50].

The main components of CNNs are convolutional layers, which allow the network to identify local conjunctions of features in input data, including EEG signals. Convolutional layers help identify complex temporal relationships between different EEG signal segments by organizing units into feature maps. They are arranged into feature maps, with each unit linked to local patches in the preceding layer's feature maps via a filter bank—a collection of shared weights. To add non-linearities to the model, the output of this local weighted sum is then run through a non-linearity, such as a Rectified Linear Unit (ReLU).

The operation of convolutional layers can be mathematically expressed as:

$$\mathrm{Y[i,j]} = \sum_{m}\sum_{n} \mathrm{X[i+m, j+n]} \, * \, \mathrm{K[m,n]} + \mathrm{b}$$

Where $\mathrm{Y[i,j]}$ is the output feature map, $\mathrm{X[i+m,j+n]}$ represents the input data, $\mathrm{K[m,n]}$ denotes the convolutional kernel, and b is the bias term. This operation is applied at every spatial location $(\mathrm{i,j})$ across the input data, resulting in the generation of the feature map [54].

In order to reduce the representation's dimensionality while keeping the most noticeable features, pooling layers are essential. Pooling layers assist in the gradual merging of semantically similar features in time series signals, like EEG data, so the network may concentrate on the most informative components of the signal. By creating invariance to slight temporal shifts and distortions, this approach strengthens the model's resistance to changes in the EEG recordings. The significant noise and severe unpredictability of the EEG signal are its distinguishing features. The algorithm's Adaptability can be somewhat increased and noise can be filtered by the pooling process [55].

After being successfully reduced in dimensionality while maintaining significant features, the processed data from the convolutional and pooling layers is subsequently passed into the fully connected layer for additional feature integration and abstraction [10]. This final stage of the CNN architecture enables the model to learn complex relationships between the extracted features and generate the desired output, such as age prediction based on EEG signals. By gradually condensing the original features through mechanisms like local receptive fields, weight sharing, and pooling, CNNs enhance the efficiency of the machine learning process and enable more accurate predictions to be made from the EEG data.

4.3.4 Deep Long Short-Term Memory (LSTM)

LSTM networks belong to a category of recurrent neural networks (RNNs) that are adept at recognizing long-term relationships and patterns within sequences of data. They excel in applications like predicting time series, processing natural language, and recognizing speech.

The architecture of an LSTM network consists of memory cells that maintain a hidden state over time, allowing them to remember information over long sequences. This is achieved through the use of specialized gating mechanisms that control the information entering and leaving each cell. The LSTM block consists of three gates: an input gate, an output gate, and a forget gate [56]. The architecture of traditional LSTM is illustrated in Figure 4-6 [57]. How much of the prior cell state to remember or forget is determined by the forget gate. The forget gate receives as inputs the current input x_t as well as the previous cell state C_{t-1} ; it then outputs a value between 0 and 1 for each cell state element, indicating how much of the prior state should be forgotten.

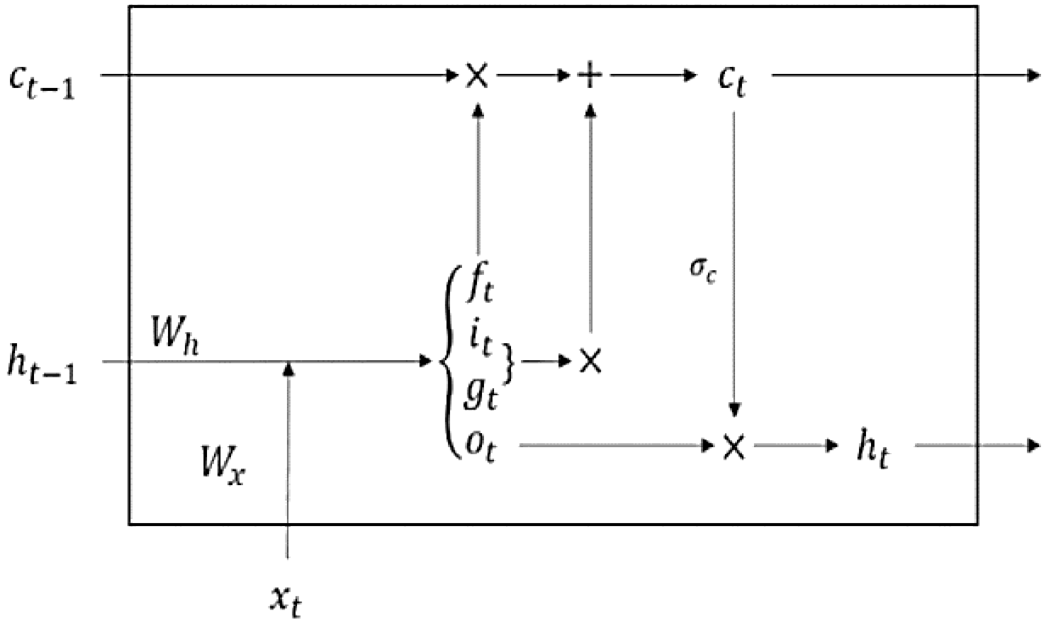

Figure 4-6. The LSTM architecture, Reprinted from "Hybrid CNN and LSTM Network For Heart Disease Prediction" by Sudha, V.K., Kumar (2023) [57], with permission from Springer Nature.

The input gate controls how much new information is added to the cell state. It takes both the previous cell state C_{t-1} and the current input x_t as input and produces a value between 0 and 1 for each element of the cell state, indicating how much of the new information to retain.

The output gate determines the output of the LSTM cell. It takes the current input x_t and the previous cell state C_{t-1} as input and produces the current cell state C_t and the current hidden state h_t, which is the output of the LSTM cell.

Mathematically, the computations performed by an LSTM cell can be described as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_t + W_{ci}C_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_t + W_{cf}C_{t-1} + b_f)$$

$$C_t = f_t.C_{t-1} + i_t.\tanh(W_{xc}x_t + W_{hc}h_t + b_c)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_t + W_{co}C_t + b_o)$$

$$h_t = o_t.\tanh(C_t)$$

Where i_t is the input gate vector, f_t is the forget gate vector, o_t is the output gate vector, C_t is the cell state, h_t is the hidden state, x_t is the input at the time step t, W represents weight matrices, b represents bias vectors, and σ is the sigmoid activation function.

4.3.5 CNN-LSTM hybrid model

In this study, we propose an approach for predicting gestational age at birth using a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures. The gestational age prediction method using a CNN-LSTM model is depicted in Figure 4-7. The suggested model is composed of one input layer, five convolutional layers with a max-pooling layer in each of them, and two LSTM layers.

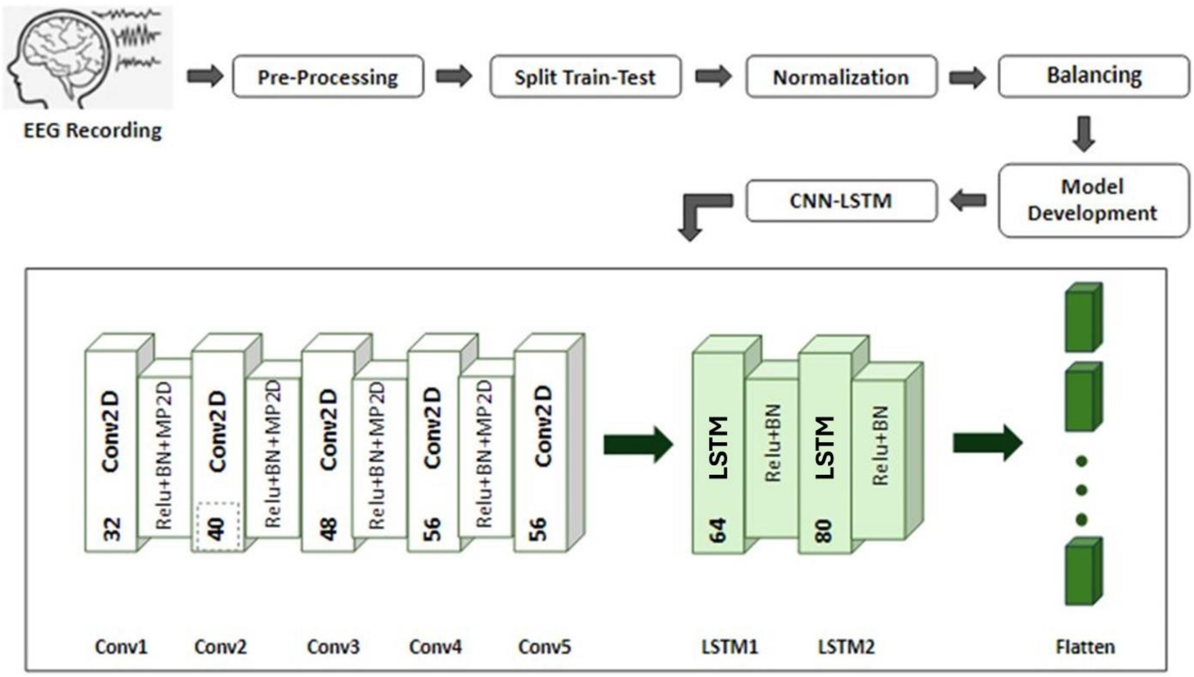

Figure 4-7. Proposed Model, Author's own drawing.

CNNs are well-suited for extracting spatial features from multi-dimensional data such as EEG signals. By applying convolutional filters across the input data, CNNs can automatically learn hierarchical representations of spatial patterns, which are then fed into subsequent layers for further processing.

LSTM networks, on the other hand, were designed specifically to capture temporal dependencies in sequential data. They are perfect for modeling temporal relationships in time-series data, such as EEG signals, because they can learn long-term dependencies and preserve memory across extended sequences. LSTM units contain a memory cell that can maintain information over time, allowing the network to learn from past observations and make predictions based on the sequence history.

By utilizing the complementing strengths of CNNs and LSTMs, our methodology is able to capture both spatial and temporal features from EEG recordings. After processing the EEG data's spatial structure, the CNN component extracts relevant spatial patterns indicative of gestational age. The CNN's output is then passed into the LSTM component, which uses the sequential dependencies over time to simulate the temporal dynamics of the EEG data. By combining these two architectures, we aim to develop a robust and accurate model for predicting gestational age at birth based on EEG signals.

4.4 Model Validation

The evaluation of the proposed CNN-LSTM hybrid model involves rigorous testing and validation procedures to assess its performance across different datasets and conditions. Central to this evaluation is the use of cross-validation, a widely adopted technique in machine learning for estimating the model's performance and generalization ability.

4.4.1 Cross-Validation

The primary focus in regression and classification is typically on the system's capacity for generalization, or how well it performs with unseen data. For the purpose of constructing an accurate estimate of this performance, the data utilised for testing is typically not used, which creates two problems. First, the difficulty of keeping testing data out of the training process, which might restrict the performance of the model, especially in cases where data is scarce. It then dives into the statistical side of things, highlighting the fact that the dataset that the researchers utilized is only one potential result of a stochastic process. As a result, every error measure that can be obtained from this dataset is only one instance of a stochastic variable along with its probability distribution. The variability and dependability of accuracy measurements are affected by this.

A method is k-fold cross-validation which is commonly used in classification and regression to address these issues, that in it all available data is randomly divided into k sets in a process. Cross-validation is a crucial technique used in machine learning to evaluate the performance of predictive models. Among its various methods, k-fold cross-validation stands out as a robust approach, particularly when dealing with limited data [58]. Then, using each set once as the test set and the remaining sets for model construction, the entire training or model fitting process is carried out k times, along with the error calculation. Consequently, all of the data is used for both training and testing after the approach eventually obtains k-independent realizations of the error measure.

K-fold cross-validation can reduce the bias and variability based on a single train-test split which is one of its main advantages. K-fold cross-validation yields a more accurate assessment of the model's generalization performance by splitting the data and testing the model on various subsets repeatedly. Additionally, it makes sure that every data point is utilized for both validation and training, making the most use of the given data. The size of the dataset and the available computing power are two important considerations for selecting k. Common choices for k are k=5 and k=10, while other numbers can also be utilized. Raising k can result in a more accurate performance estimate, but it also raises the cost of computing.

In our approach, we recognize the importance of avoiding overfitting, a frequent challenge in machine learning where a model performs admirably on training data but struggles to apply its findings to unseen data. To mitigate this risk, we initially split the available data into training and testing sets and held out test set to evaluate the final model's performance. Subsequently, we adopted a cross-validation strategy to further evaluate the robustness and generalization capability of our proposed CNN-LSTM architecture for age prediction which is presented in Figure 4-8.

To conduct cross-validation, we split the training data into five folds, ensuring each fold represented a diverse subset of samples, as illustrated in Figure 4-9 [59]. Within each iteration of the cross-validation loop, one fold was reserved as the validation set, while the remaining folds were used for training. This approach allowed us to thoroughly assess the model's performance across different subsets of the training data.

By integrating cross-validation, our goal was to enhance the reliability and robustness of our deep learning prediction model, ensuring its performance could be assessed across different datasets and minimizing the risk of overfitting to any specific training set. Our main goal with cross-validation was to verify how well the model performs on data it hasn't seen before and to evaluate its capacity to work with new samples, rather than focusing on choosing the best hyperparameters. This iterative process allowed us to confidently evaluate the effectiveness of our model across various datasets and scenarios.

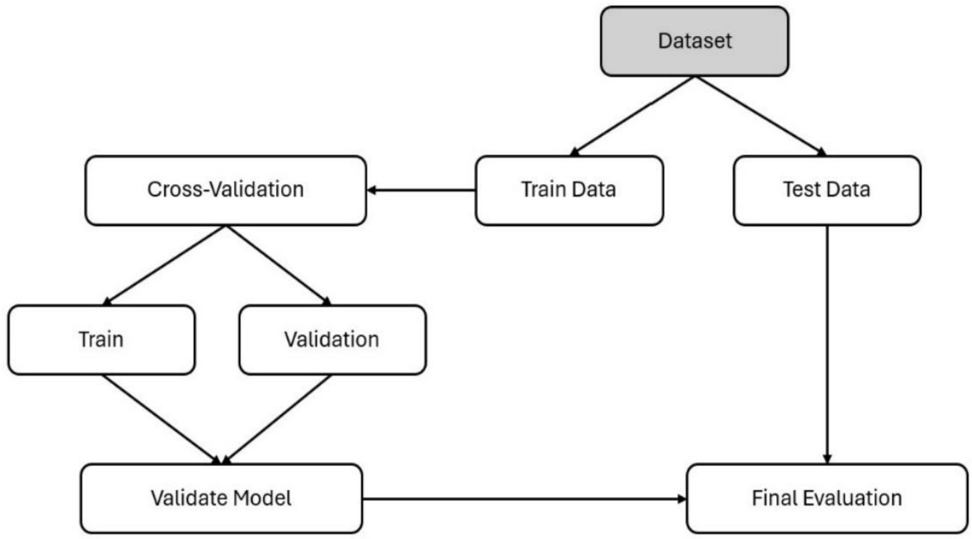

Figure 4-8. Cross-validation flowchart, Author's own drawing.

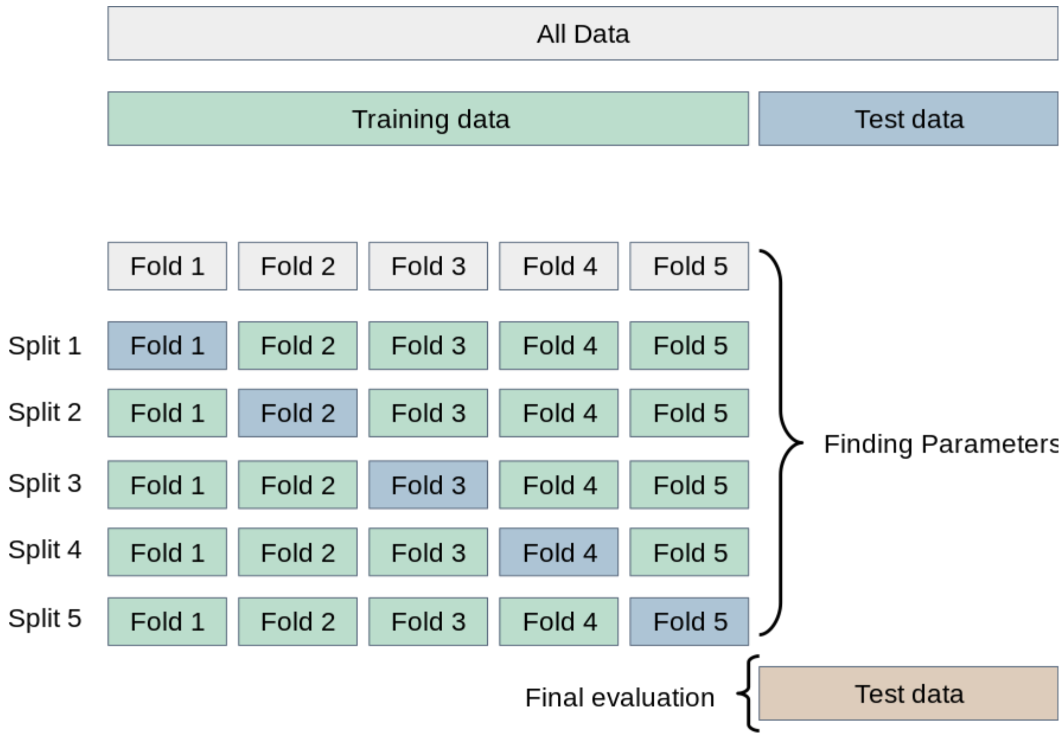

Figure 4-9. 5-Fold Cross-validation, Reprinted from "Infrared Thermal Imaging-Based Turbine Blade Crack Classification Using Deep Learning" by Benedict E. Jaeger, Simon Schmid, Christian U. Grosse, Anian Gögelein, and Frederik Elischberger. (2022) [59], which is licensed under a Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).

4.4.2 Regularization

Regularization is a fundamental technique in machine learning, designed to reduce model complexity while maintaining predictive accuracy on unseen data. This is particularly crucial in the context of deep learning, where the highly nonlinear nature of deep neural networks can lead to significant overfitting if not properly managed. This approach is especially vital in deep learning due to the deep neural networks' highly nonlinear behaviors. The nonlinearity of deep networks allows them to capture complex patterns, but it also increases the risk of overfitting on the training data as more layers are added [60].

Deep neural networks, due to their large number of parameters, are highly prone to overfitting. Overfitting happens when a model picks up on the noise in the training data rather than the actual patterns, which results in subpar performance when tested with new data [61]. Regularization methods are therefore essential in protecting these models from overfitting, ensuring they maintain robust performance across both training and unseen data. To address this challenge, we employ early stopping and L2 regularization.

4.4.2.1 Early Stopping

Early stopping is a method used to stop training a model when its performance on a validation set starts to decline. This approach effectively prevents the model from becoming overly complex and provides a form of implicit regularization [62]. Effective early stopping involves monitoring a metric on a separate validation set to decide the optimal point to halt training, thus mitigating the risk of overfitting. For robust networks, early stopping is particularly advantageous, as it helps achieve the best possible performance even during adversarial training. By monitoring the validation error during training and stopping when it starts to rise, early stopping ensures that the model generalizes well to unseen examples [63].

In the context of neural network training, early stopping serves as a powerful tool to combat overfitting and improve generalization performance. By terminating the training process at the right moment, early stopping prevents the model from fitting noise in the training data and encourages it to capture meaningful patterns instead [64]. This is particularly useful when dealing with complex models that have a high capacity to memorize the training data, which can lead to poor performance on unseen data.

In our model, we incorporated early stopping to enhance generalization and prevent overfitting. During each fold of cross-validation, we monitored the validation loss during training and used an early stopping callback with a patience of 10 epochs. This implies that if there was no improvement in the validation loss for 10 consecutive epochs, the training was stopped, and the model's parameters were reverted to the state where the validation loss was the lowest. This strategy allowed us to halt training before the model began to overfit the training data, ensuring that the learned patterns were robust and generalizable to new data.

By implementing early stopping, we observed that our model's performance on the validation set improved significantly. The validation loss stabilized and converged, indicating that the model was no longer overfitting the training data. This approach resulted in more reliable and interpretable predictions, as evidenced by the reduced discrepancy between training and validation metrics. Overall, early stopping was a crucial component of our training process, helping to achieve a balance between model complexity and generalization ability.

After completing the cross-validation process, we selected the final model based on the performance metrics obtained from the cross-validation folds. We then trained this final model on the entire training set (excluding the held-out test set) using the same early stopping criteria. The held-out test set, which was not used during the training or cross-validation process, was subsequently used to evaluate the final model's performance independently. This approach provided an unbiased assessment of the model's effectiveness in predicting age from EEG signals, ensuring that our findings were robust and generalizable to unseen data.

4.4.2.2 L2 Regularization

L2 regularization, also known as weight decay, is a common technique used to prevent overfitting in machine learning models, particularly in deep neural networks. It operates by adding a penalty term to the loss function, proportional to the sum of the squared weights of the model [60].

The primary goal of L2 regularization is to encourage the model to learn simpler patterns by penalizing large weight values. This regularization term effectively imposes a constraint on weights, discouraging them from growing too large during the training process. As a result, the model becomes less affected by minor variations in the training data, which enhances its ability to perform well on new, unseen data. This regularization technique promotes smoother decision boundaries and reduces the model's reliance on individual features, resulting in more stable and reliable predictions across different datasets [65].

In our deep learning model architecture, we employed L2 regularization to mitigate the risk of overfitting. Specifically, L2 regularization was applied to the convolutional and LSTM layers of the model. By incorporating L2 regularization into these layers, we aimed to impose constraints on the weights of the neural network, discouraging excessively large weight values. This regularization technique helps in preventing the model from overly fitting to irrelevant details in the training data, thereby promoting more robust generalization to unseen data. Moreover, L2 regularization discourages the model from emphasizing large weights, thereby encouraging it to learn simpler representations that are more resistant to overfitting. Overall, the integration of L2 regularization into our model architecture enhances its ability to learn meaningful patterns from the data while ensuring that the learned representations are more generalizable and reliable.

4.5 Assessment of Prediction Performance

By iterative training and evaluating the model on different subsets of the data, we obtain reliable estimates of its performance metrics, such as Mean Absolute Error, Root Mean

Square Error, and Coefficient of determination scores. This rigorous cross-validation process allows us to assess the model's predictive accuracy, variability, and generalization capability, providing valuable insights for model refinement and validation.

4.5.1 Mean Absolute Error (MAE)

MAE serves as a valuable metric for assessing the performance of regression models. Unlike RMSE, which increases the impact of greater errors because of the squaring operation, MAE considers all errors equally, offering a straightforward and intuitive measure of error magnitude. Regardless of the direction or size of the deviation, MAE considers each error's absolute value, ensuring that both small and large errors have an equal impact on the overall assessment. By averaging these absolute errors, MAE provides insight into the average magnitude of deviations between predicted and actual values. This characteristic makes MAE particularly useful when you want to gauge overall model accuracy without disproportionately penalizing outliers or large errors [66].

The MAE can be mathematically expressed as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i|$$

Where in all the above equations Y_i is the actual ith value, and X_i represents the predicted ith value. Also m is the number of observations.

4.5.2 Root Mean Square Error (RMSE)

RMSE represents the square root of the mean of the squared errors, making it a widely adopted metric in evaluating numerical predictions. It provides a measure of the typical magnitude of errors in the model's predictions, with a greater emphasis on larger errors. The interpretation of RMSE is in the same units as the response variable, facilitating a direct comparison with the predicted variable. A lower RMSE value indicates a better fit to the data, whereas a higher value suggests a poorer fit. However, achieving an RMSE of zero or an extremely low value may indicate overfitting to the training data, leading to poor generalization on unseen data. This principle is generally true for many performance metrics, including MAE and Coefficient of determination, where a perfect score might suggest that the model has learned noise from the training data rather than the underlying pattern [66]. It is calculated as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2}$$

4.5.3 Coefficient of determination (R^2)

R^2, or the coefficient of determination is a fundamental metric used to assess the quality of a regression model. It quantifies the extent to which the observed outcomes are accounted for by the model, representing the proportion of the total variability in the dependent variable that can be explained by the independent variable(s). In essence, R^2 signifies the predictability of the dependent variable based on the independent variable(s), with a value of 100% denoting a complete explanation of variability. This metric, often regarded as one of the most prominent in regression analysis, offers insight into how well the model captures the underlying patterns and relationships within the data, thus aiding in model evaluation and interpretation [66]. It can be computed mathematically as:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2}$$

5 Results and Discussion

Our investigation delves into the outcomes yielded by our developed model for predicting gestational age at birth, juxtaposing its performance against traditional CNN and LSTM architectures, as well as a hybrid CNN-LSTM model. Through an in-depth analysis of various evaluation metrics, we aim to discern the effectiveness of our proposed approach in accurate gestational age estimation. By examining how well our model generalizes to unseen data, we can ascertain its robustness and potential for real-world applications. Furthermore, the comparison comparative analysis with CNN and LSTM models allows us to identify the strengths and weaknesses of different model architectures and gauge the relative performance of our proposed approach.

Additionally, we broaden our perspective by comparing our results with those reported in existing literature on gestational age prediction. By contextualizing our findings within the broader research landscape, we can validate the significance and novelty of our contributions while identifying avenues for further improvement and exploration. Through this comprehensive evaluation, we aim to provide insights that advance the field of gestational age prediction and pave the way for more accurate and reliable clinical practices.

Embedded within this framework is a detailed exploration of training our proposed CNN-LSTM hybrid model for age prediction using EEG data. Model training is pivotal in developing predictive models, especially in deep learning. Through a robust training procedure involving cross-validation, we ensured reliable and unbiased model selection. The train set was divided into five folds to iteratively train the model on different data subsets, thereby allowing for more reliable model selection and assessment of its generalization capabilities. Preprocessing techniques such as Z-score normalization and data balancing were employed to ensure data consistency and mitigate class imbalances. Utilizing an MSE loss function and Adam optimizer, we optimized the model parameters while monitoring key performance metrics like MAE, RMSE, and R^2 to assess convergence and generalization.

In the sections that follow, we go into greater detail on the results of our model training process. We examine how well the model performed on training and test datasets, contrast it with other deep learning models and place our findings in the context of the body of the existing literature.

5.1 Experimental Results

5.1.1 Prediction Performance on Training and Test Set

The evaluation of our gestational age prediction model's performance on both the training and test sets provides valuable insights into its robustness and generalization capabilities, as illustrated in Table 5-1. Our model exhibited commendable performance on the training set, as evidenced by low MAE and RMSE values of 3.01 days and 4.12 days, respectively. These results indicate that the model accurately predicted gestational age when trained on the available data.

Moving to the test set, which comprises unseen data, our model maintained competitive performance, albeit with slightly higher MAE and RMSE values of 3.16 days and 4.38 days, respectively. Despite these marginal increases, the model's ability to maintain accuracy on the test set underscores its capacity to generalize well to new samples, a crucial aspect in real-world applications. The consistency in performance metrics between the training and test sets suggests that our model effectively captured underlying patterns in the data without succumbing to overfitting, thereby ensuring reliable predictions for diverse patient populations.

Table 5-1. Performance on Training and Test set

Data	MAE (days)	RMSE	R^2
Train Set	3.01	4.12	0.86
Test Set	3.16	4.38	0.75

Furthermore, the R^2 values of 0.86 on the training set and 0.75 on the test set indicate a strong correlation between predicted and actual gestational ages. These high R^2 values signify that our model accounts for a significant portion of the variability in gestational age, further validating its efficacy in inaccurate age predictions.

The Figure 5-1 illustrates the performance of the regression model in predicting ages. Each blue dot represents a data point, showing the predicted age against the actual gestational age. The black dashed line represents the ideal case where the predicted age is equal to the real age. Points lying close to the dashed line indicate accurate predictions, whereas points farther away indicate larger prediction errors. This visualization helps in assessing the model's accuracy and the overall distribution of prediction errors. It can be observed that the predicted values are generally close to the ideal case, indicating that our model has been trained well.

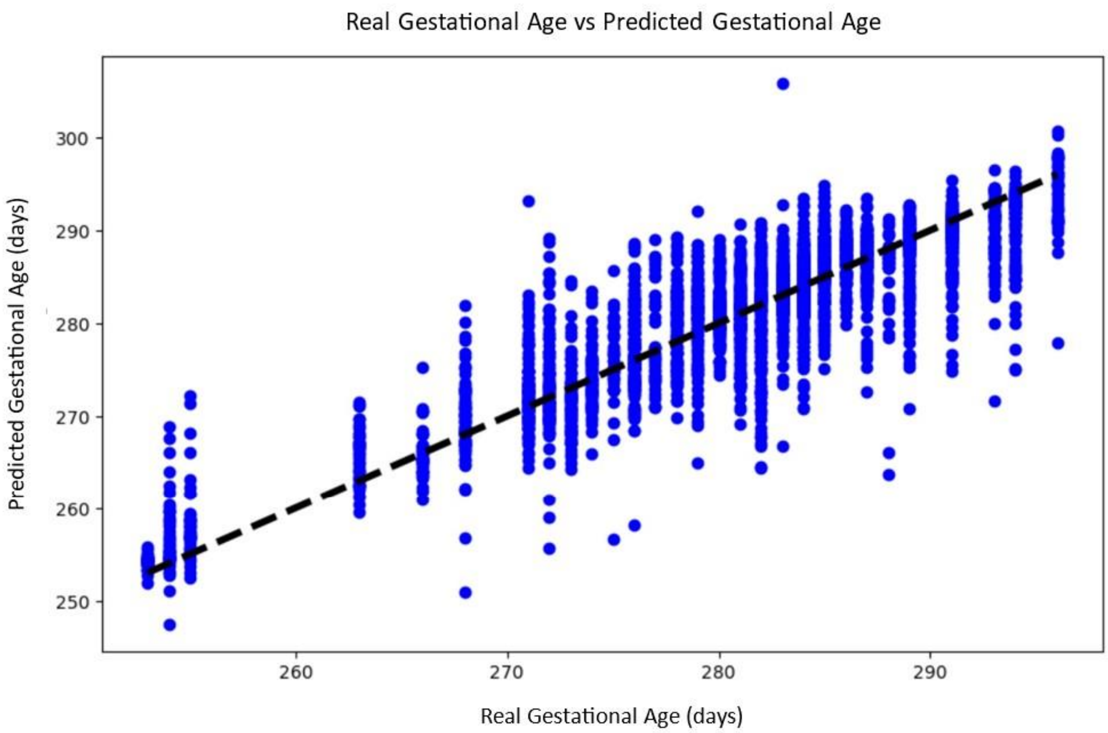

Figure 5-1. Scatter plot of predicted gestational ages versus real gestational ages,

In summary, our evaluation demonstrates that our gestational age prediction model performs well on both the training and test sets, showcasing its reliability, generalization ability, and capacity to provide accurate predictions for diverse patient populations. These findings bolster confidence in the utility of our model as a valuable tool in clinical settings for estimating gestational age and informing clinical decision-making processes.

5.1.2 Prediction Performance Compared with Other Deep Learning Models

In comparing the prediction performance of our proposed hybrid model with traditional CNN and LSTM models, we observed significant differences in their ability to accurately predict gestational age. While CNN and LSTM models are widely used in various time-series prediction tasks, they cannot inherently effectively capture the temporal dependencies and spatial features present in gestational age data. CNN models excel at extracting spatial features from two-dimensional data but may struggle with sequential data such as time series. Similarly, LSTM models are adept at capturing long-range dependencies in sequential data but may not effectively leverage spatial information.

Our proposed hybrid model integrates the strengths of both CNN and LSTM architectures, thereby enhancing its ability to capture both temporal and spatial features inherent in gestational age data. By leveraging the hierarchical feature extraction capabilities of CNN layers followed by the sequential learning ability of LSTM layers, our hybrid model can effectively model complex patterns and relationships in gestational age data, leading to more accurate predictions.

Table 5-2. Performance of Other Deep Learning Models

Model	Train MAE (Days)	Test MAE (Days)
CNN	12	8.9
LSTM	9.17	7.39
CNN-LSTM hybrid model	3.09	3.16

Table 5-2 showcases the performance comparison of various deep learning models that we used in predicting gestational age, highlighting the MAE metric. The CNN model yielded an MAE of12 for train data and 8.9 for test data, while the LSTM model performed slightly better with an MAE of 9.17 for train data and 7.39 for test data. In contrast, our CNN-LSTM hybrid model significantly outperformed both standalone models in both train and test, achieving an impressive MAE of 3.09 and 3.16 for train and test data.

Through rigorous experimentation and analysis, we anticipate that our hybrid model will continue to demonstrate superior performance over standalone CNN and LSTM models in predicting gestational age. One of the key advantages of our hybrid approach is its ability to leverage both spatial and temporal information, allowing for a more comprehensive understanding of the underlying patterns in gestational age data. However, it's important to note that each model architecture has its own set of advantages and disadvantages. While CNN models excel at extracting spatial features, they may struggle with capturing temporal dependencies. Conversely, LSTM models are proficient at capturing temporal dependencies but may overlook spatial features. Our hybrid model mitigates these limitations by combining the strengths of both architectures, offering a more robust and accurate solution for gestational age prediction in clinical settings.

5.1.3 Prediction Performance Compared with Other Literature

In comparing our proposed hybrid model's prediction performance with existing literature, we assessed three notable studies in the field of neonatal age prediction using machine learning techniques.

Ansari et al. [45] conducted a study aiming to estimate neonates' biological brain age utilizing a deep learning model trained on resting-state EEG data. Their model achieved an MAE of 1.03 weeks and 0.98 weeks in two independent datasets, demonstrating proficiency in discerning brain age gaps between neonates with normal and severely abnormal outcomes [45]. Similarly, another study by Ansari et al. [46] proposed a deep-learning approach for predicting brain age in preterm neonates using EEG data. Leveraging a CNN block based on the Inception architecture, their model achieved an impressive MAE of 0.78 weeks, showcasing its ability to differentiate between neonates with normal and severely abnormal outcomes [46].

In addition to these studies, we examined the research conducted by Stevenson et al. [41]. Their objective was to overcome the effects of site differences in EEG-based brain age prediction in preterm neonates. Utilizing a 'bag of features' approach with a combination function estimated using SVR and feature selection, they aimed to predict post-menstrual age from EEG recordings. Notably, their study reported an MAE of 1.0 weeks when training the age predictor on data from one site, which improved to 1.1 weeks when applied to independent data from another site. This improvement in validation accuracy was achieved with a reduced feature set, demonstrating the importance of feature selection in enhancing prediction performance [41].

In summary, while our proposed hybrid model demonstrates promising results in predicting gestational age, it is essential to consider and compare its performance with existing literature to gain a comprehensive understanding of its efficacy and potential applications in clinical settings. The comparison Table 5-3 provides quantitative evidence of the relative performance of our proposed hybrid model compared to other machine learning approaches.

Table 5-3- Performance Compared with Other Literature

Article	**MAE (Days)**
Our hybrid model (CNN-LSTM)	3.16
A deep learning model trained on resting-state EEG data [45]	6.86-7.21
Leveraging a CNN block based on the Inception architecture (Sinc), [46]	5.46
Inter-site generalizability of EEG based age prediction algorithms in the preterm infant [41]	4.9

5.2 Discussion

Our study used the hybrid CNN-LSTM architecture with EEG data to develop and evaluate a model for prediction in determining gestational age at birth. We could draw valuable insights into the functionality and possible uses of our method from the careful examination of our findings and the comparison with previous models and research.

In the evaluation of the prediction performance on the test set, our model showed commendable results so, the low MAE and RMSE values mean our model can estimate the gestational age precisely. Notably, the model showed competitive performance on the unseen test set, underlining its robustness and generalization capabilities. Consistency in performance metrics across different subsets of data corroborates the fact that our model captured underlying patterns well, without succumbing to overfitting.

Comparing our hybrid model with traditional CNN and LSTM architectures revealed significant differences in prediction performance. While CNN and LSTM models have their respective strengths, our hybrid approach integrates the benefits of both architectures, enabling more accurate and comprehensive gestational age prediction. The superior performance of our hybrid model, as evidenced by the significantly lower MAE, underscores its potential utility in clinical settings.

In addition, a comparative analysis with previous research was conducted to provide context and corroborate the results. The efficacy of our hybrid model in predicting neonate gestational age using EEG data contrasted positively with earlier investigations.

In conclusion, our study presents a novel approach to gestational age prediction using EEG data and a hybrid CNN-LSTM architecture. The robust performance of our model, coupled with its comparative advantages over traditional CNN and LSTM models, underscores its potential as a valuable tool for clinicians in estimating gestational age and informing clinical decision-making. Moving forward, continued research and validation efforts are essential to further refine and optimize our predictive model for real-world applications in neonatal care.

6 Conclusion

In this work, we used EEG data to develop and assess a hybrid CNN-LSTM model to predict gestational age at birth. Our investigation attempted to overcome the limitations of conventional CNN and LSTM architectures by combining their advantages to provide gestational age prediction that is more accurate and trustworthy. Through a comprehensive analysis of our model's performance and comparisons with existing methods, we have gained valuable insights into its potential applications and limitations.

Our model exhibited promising results in predicting gestational age, outperforming traditional CNN and LSTM models in terms of accuracy and robustness. The integration of both spatial and temporal features through the hybrid architecture allowed for more comprehensive modeling of gestational age data, leading to improved prediction performance. The competitive performance of our model, as evidenced by low MAE (MAE=3.16 days) and RMSE (RMSE=4.38 days) values, underscores its potential utility in clinical settings for estimating gestational age and informing clinical decision-making processes.

While our study provides promising results, several limitations and areas for future research should be acknowledged. Firstly, the dataset used for model training and evaluation may have inherent biases or limitations, potentially affecting the generalizability of our findings. Additionally, the performance of our model may vary across different demographic or clinical settings, warranting further validation and refinement. Future studies could explore the integration of additional data modalities or features to enhance prediction accuracy and robustness. Moreover, the time-consuming nature of deep learning models and their black-box nature may limit the interpretability of our model for clinicians and in multi-disciplinary environments. Last but not least, the neonates in our study were healthy and birth after gestational week 36, and the model's performance may vary in populations with different health statuses or preterm neonates. Next, the dynamic nature of neonatal EEG data due to ongoing brain development poses challenges in capturing meaningful information for accurate gestational age prediction.

The promising results, evidenced by low MAE and RMSE values, suggest the model's potential utility in clinical settings for estimating gestational age and supporting clinical decision-making. However, further validation and refinement are needed to ensure its generalizability and robustness across various clinical and demographic settings.

In conclusion, while our study presents a promising approach to gestational age prediction using EEG data and a hybrid CNN-LSTM architecture, there are important considerations and areas for future research. Further validation and refinement of our model are warranted to enhance its generalizability and robustness across different demographic and clinical settings. Future studies could explore the integration of additional data modalities or features to improve prediction accuracy and address the limitations identified in this study. Ultimately, the development of more accurate and reliable predictive models for gestational age estimation holds great promise for improving clinical outcomes and patient care in neonatal medicine.

7 Future works

In this chapter, we outline several avenues for future research to enhance the capabilities and applications of gestational age prediction models using EEG data. These directions aim to address current limitations, improve model performance, and expand the utility of the developed models in diverse clinical settings.

7.1 Feature Selection and Optimization

One promising area for future research is the optimization of feature selection in EEG data. Currently, our model processes data from multiple EEG channels, some of which may not significantly contribute to the prediction of gestational age. By investigating the correlation between different EEG channels, future studies could identify and eliminate those that provide redundant or insignificant information. This refined feature selection process would not only streamline the computational requirements but also enhance the model's performance. Employing techniques like principal component analysis (PCA), independent component analysis (ICA), or other sophisticated dimensionality reduction methods can lead to the development a more efficient and accurate model.

7.2 Generalizability Across Demographics

Ensuring the robustness and generalizability of the model across diverse demographic groups and clinical settings is another critical area for future research. The current study primarily focuses on a specific population, which may introduce biases and limit the model's applicability. Future research should aim to validate and refine the model using data from various demographic groups, including different ethnicities, age groups, and clinical conditions. This would involve collecting and integrating large, diverse datasets and employing cross-validation techniques to ensure the model performs consistently well across different populations. Such efforts would enhance the model's reliability and facilitate its adoption in a wide range of clinical environments.

7.3 Improving Model Interpretability

A significant challenge with deep learning models, including the CNN-LSTM hybrid used in this study, is their black-box nature, which limits interpretability. For clinicians to trust and effectively utilize these models, it is essential to develop methods that make the predictions more transparent and understandable. Future work should focus on creating techniques that can explain the model's decision-making process. Approaches that could be explored to highlight which features or EEG channels are most influential in the model's predictions. Enhancing interpretability will bridge the gap between complex models and clinical practice, fostering greater acceptance and utilization of these advanced tools.

7.4 Incorporating Preterm Neonate Data

Extending the study to include data from preterm neonates represents another vital direction for future research. Preterm infants have different health conditions and developmental trajectories compared to full-term infants, and it is crucial to evaluate the model's performance in these cases. By incorporating preterm neonate data, researchers can assess whether the model can accurately predict gestational age across a broader spectrum of developmental stages. This expansion would also involve developing strategies to handle the unique challenges posed by preterm EEG data, such as higher variability and distinct patterns of brain activity.

7.5 Real-Time Application Development

The ultimate goal of this research is to develop practical tools that can be seamlessly integrated into clinical workflows. Future work should focus on creating a user-friendly software tool for real-time gestational age prediction. This tool would leverage the trained deep learning model to provide instant predictions based on EEG data collected in clinical settings. It should be designed with an intuitive interface, allowing clinicians to input EEG data and receive immediate, actionable insights. Integration with existing clinical systems and ensuring compliance with healthcare standards and regulations would be essential steps in this process. Additionally, real-time feedback and continuous learning mechanisms could be incorporated to update the model with new data, ensuring its accuracy and relevance over time.

References

[1] C. D. Lynch and J. Zhang, "The research implications of the selection of a gestational age estimation method," *Paediatr Perinat Epidemiol*, vol. 21 Suppl 2, pp. 86–96, Sep. 2007, doi: 10.1111/j.1365-3016.2007.00865.x.

[2] K. K. Ryckman, S. L. Berberich, and J. M. Dagle, "Predicting gestational age using neonatal metabolic markers," *American Journal of Obstetrics and Gynecology*, vol. 214, no. 4, p. 515.e1-515.e13, Apr. 2016, doi: 10.1016/j.ajog.2015.11.028.

[3] S. Sazawal, K.K.Ryckman, S.Das, R.Khanam, I.Nisar, E.Jasper, A.Dutta, S.Rahman, U.Mehmood, B.Bedell, S.Deb, N.Haque Chowdhury, A.Barkat, H.Mittal, S.Ahmed, F.Khalid, R.Raqib, A.Manu, S.Yoshida, M.Ilyas, A.Nizar, S.Mohammed Ali, A.H. Baqui, F.Jehan, U.Dhingra, and R.Bahl, "Machine learning guided postnatal gestational age assessment using new-born screening metabolomic data in South Asia and sub-Saharan Africa," *BMC Pregnancy and Childbirth*, vol. 21, no. 1, p. 609, Sep. 2021, doi: 10.1186/s12884-021-04067-y.

[4] M. Saeidi, W.Karwowski, F.V. Farahani, K.Fiok, R.Taiar, P. A. Hancock, and A.Al-Juaid, "Neural Decoding of EEG Signals with Machine Learning: A Systematic Review," *Brain Sciences*, vol. 11, no. 11, Art. no. 11, Nov. 2021, doi: 10.3390/brainsci11111525.

[5] K. Jusseaume and I. Valova, "Brain Age Prediction/Classification through Recurrent Deep Learning with Electroencephalogram Recordings of Seizure Subjects," *Sensors*, vol. 22, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/s22218112.

[6] G. Sinha, N.Belacel, Z.Gu, S.M Doesburg, G.Medvedev, U.Ribary, V.A Vakorin, P.Xi, "Machine Learning Methods for Electroencephalogram-Based Age Prediction," in *2023 IEEE SENSORS*, Oct. 2023, pp. 1–4. doi: 10.1109/SENSORS56945.2023.10324895.

[7] L. E. Ismail and W. Karwowski, "A Graph Theory-Based Modeling of Functional Brain Connectivity Based on EEG: A Systematic Review in the Context of Neuroergonomics," *IEEE Access*, vol. 8, pp. 155103–155135, 2020, doi: 10.1109/ACCESS.2020.3018995.

[8] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The PREP pipeline: standardized preprocessing for large-scale EEG analysis," *Front Neuroinform*, vol. 9, p. 16, 2015, doi: 10.3389/fninf.2015.00016.

[9] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *Journal of Neuroscience Methods*, vol. 346, p. 108885, Dec. 2020, doi: 10.1016/j.jneumeth.2020.108885.

[10] A. Liu, J. Li, and H. Ye, "A Prediction Model Combining Convolutional Neural Network and LSTM Neural Network," in *2023 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS)*, Jul. 2023, pp. 318–321. doi: 10.1109/AIARS59518.2023.00071.

[11] Shams Forruque Ahmed, Md. Sakib Bin Alam, Maruf Hassan, Mahtabin Rodela Rozbu, Taoseef Ishtiak, Nazifa Rafa, M. Mofijur, A. B. M. Shawkat Ali, and Amir H. Gandomi, "Deep

learning modelling techniques: current progress, applications, advantages, and challenges," *Artif Intell Rev*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, doi: 10.1007/s10462-023-10466-8.

[12] F. Siddiqui, A.Mohammad, M. Afshar Alam, S.Naaz, P.Agarwal, S.Saquib Sohail, and D.Øivind Madsen, "Deep Neural Network for EEG Signal-Based Subject-Independent Imaginary Mental Task Classification," *Diagnostics (Basel)*, vol. 13, no. 4, p. 640, Feb. 2023, doi: 10.3390/diagnostics13040640.

[13] A. Biasiucci, B. Franceschiello, and M. M. Murray, "Electroencephalography," *Current Biology*, vol. 29, no. 3, pp. R80–R85, Feb. 2019, doi: 10.1016/j.cub.2018.11.052.

[14] G. R. Müller-Putz, "Chapter 18 - Electroencephalography," in *Handbook of Clinical Neurology*, vol. 168, N. F. Ramsey and J. del R. Millán, Eds., in Brain-Computer Interfaces, vol. 168. , Elsevier, 2020, pp. 249–262. doi: 10.1016/B978-0-444-63934-9.00018-4.

[15] S. Siuly, Y. Li, and Y. Zhang, *EEG Signal Analysis and Classification*. in Health Information Science. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-47653-7.

[16] H. Cho and J. Paik, "Well-Being of Decolonizing Aesthetics: New Environment of Art with BCI in HCI," in *Human Interface and the Management of Information: Applications and Services*, S. Yamamoto, Ed., Cham: Springer International Publishing, 2016, pp. 244–255. doi: 10.1007/978-3-319-40397-7_24.

[17] A. J. Casson, D. C. Yates, S. J. M. Smith, J. S. Duncan, and E. Rodriguez-Villegas, "Wearable Electroencephalography," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 3, pp. 44–56, May 2010, doi: 10.1109/MEMB.2010.936545.

[18] K. A. Carden, "Recording Sleep: The Electrodes, 10/20 Recording System, and Sleep System Specifications," *Sleep Medicine Clinics*, vol. 4, no. 3, pp. 333–341, Sep. 2009, doi: 10.1016/j.jsmc.2009.04.002.

[19] G. Klem, H. Lüders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology.," *Electroencephalography and clinical neurophysiology. Supplement*, 1999, Accessed: May 21, 2024. [Online]. Available: https://www.semanticscholar.org/paper/The-ten-twenty-electrode-system-of-the-Federation.-Klem-L%C3%BCders/53a7cf6bf8568c660240c080125e55836d507098

[20] M. S. Scher, "Neurophysiological assessment of brain function and maturation. II. A measure of brain dysmaturity in healthy preterm neonates," *Pediatr Neurol*, vol. 16, no. 4, pp. 287–295, May 1997, doi: 10.1016/s0887-8994(96)00009-4.

[21] M. André, M. -D. Lamblin, A. M. d'Allest, L. Curzi-Dascalova, F. Moussalli-Salefranque, S. Nguyen The Tich, M. -F. Vecchierini-Blineau, F. Wallois, E. Walls-Esquivel, and P. Plouin, "Electroencephalography in premature and full-term infants. Developmental features and glossary," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 40, no. 2, pp. 59–124, May 2010, doi: 10.1016/j.neucli.2010.02.002.

[22] L. Webb, M. Kauppila, J. A. Roberts, S. Vanhatalo, and N. J. Stevenson, "Automated detection of artefacts in neonatal EEG with residual neural networks," *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106194, Sep. 2021, doi: 10.1016/j.cmpb.2021.106194.

[23] N. Stevenson and A. Tokariev, "Automated EEG Analysis for Neonatal Intensive Care," in *Encyclopedia of Biomedical Engineering*, R. Narayan, Ed., Oxford: Elsevier, 2019, pp. 240–257. doi: 10.1016/B978-0-12-801238-3.10889-X.

[24] E. Shany and I. Berger, "Neonatal Electroencephalography: Review of a Practical Approach," *J Child Neurol*, vol. 26, no. 3, pp. 341–355, Mar. 2011, doi: 10.1177/0883073810384866.

[25] Laura Anna van de Pol, Charlotte van 't Westende, Inge Zonnenberg, Esther Koedam, Ineke van Rossum, Willem de Haan, Marjan Steenweg, Elisabeth Catharina van Straaten, and Cornelis Jan Stam, "Strong Relation Between an EEG Functional Connectivity Measure and Postmenstrual Age: A New Potential Tool for Measuring Neonatal Brain Maturation," *Front. Hum. Neurosci.*, vol. 12, Jul. 2018, doi: 10.3389/fnhum.2018.00286.

[26] E.Cohen, Y. Wong, Euan M. Wallace, Joanne C. Mockler, Alexsandria Odoi, Samantha Hollis, Rosemary S. C. Horne, and Stephanie R. Yiallourou, "EEG power spectrum maturation in preterm fetal growth restricted infants," *Brain Research*, vol. 1678, pp. 180–186, Jan. 2018, doi: 10.1016/j.brainres.2017.10.010.

[27] K. Blinowska and P. Durka, "Electroencephalography (EEG)," in *Wiley Encyclopedia of Biomedical Engineering*, John Wiley & Sons, Ltd, 2006. doi: 10.1002/9780471740360.ebs0418.

[28] D. P. Subha, P. K. Joseph, R. Acharya U, and C. M. Lim, "EEG Signal Analysis: A Survey," *J Med Syst*, vol. 34, no. 2, pp. 195–212, Apr. 2010, doi: 10.1007/s10916-008-9231-z.

[29] S. Das, A. Dey, A. Pal, and N. Roy, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect," *IJCA*, vol. 115, no. 9, pp. 31–41, Apr. 2015, doi: 10.5120/20182-2402.

[30] M. s. Roobini, M. Sowmiya, S. Jancy, and L. Suji Helen, "Applications of Machine Learning Techniques in Disease Detection," in *Medical Imaging and Health Informatics*, John Wiley & Sons, Ltd, 2022, pp. 23–45. doi: 10.1002/9781119819165.ch2.

[31] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.

[32] A. Fluck, R. Trippe, L. Onderwijs, and J. Magenheim, "State of the art in thinking about machine learning – implications for education," 2019.

[33] G. Li, C. H. Lee, J. J. Jung, Y. C. Youn, and D. Camacho, "Deep learning for EEG data analytics: A survey," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 18, p. e5199, 2020, doi: 10.1002/cpe.5199.

[34] O. Al Zoubi , Chung Ki Wong, Rayus T. Kuplicki, Hung-wen Yeh, Ahmad Mayeli, Hazem Refai, Martin Paulus, and Jerzy Bodurka, "Predicting Age From Brain EEG Signals—A Machine

Learning Approach," *Frontiers in Aging Neuroscience*, vol. 10, 2018, Accessed: Feb. 01, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2018.00184

[35] J. M. O'Toole, G. B. Boylan, S. Vanhatalo, and N. J. Stevenson, "Estimating functional brain maturity in very and extremely preterm neonates using automated analysis of the electroencephalogram," *Clinical Neurophysiology*, vol. 127, no. 8, pp. 2910–2918, Aug. 2016, doi: 10.1016/j.clinph.2016.02.024.

[36] N. J. Stevenson , L. Oberdorfer, N. Koolen, J. M. O'Toole, T. Werther, K. Klebermass-Schrehof, and S. Vanhatalo, "Functional maturation in preterm infants measured by serial recording of cortical activity," *Sci Rep*, vol. 7, no. 1, Art. no. 1, Oct. 2017, doi: 10.1038/s41598-017-13537-3.

[37] S. I. Dimitriadis and C. I. Salis, "Mining Time-Resolved Functional Brain Graphs to an EEG-Based Chronnectomic Brain Aged Index (CBAI)," *Front Hum Neurosci*, vol. 11, p. 423, 2017, doi: 10.3389/fnhum.2017.00423.

[38] H. Sun, Luis Paixao, Jefferson T. Oliva, Balaji Goparaju, Diego Z. Carvalho, Kicky G. van Leeuwen, Oluwaseun Akeju, Robert J. Thomas, Sydney S. Cash, Matt T. Bianchi, and M. Brandon Westover, "Brain age from the electroencephalogram of sleep," *Neurobiology of Aging*, vol. 74, pp. 112–120, Feb. 2019, doi: 10.1016/j.neurobiolaging.2018.10.016.

[39] M. Klymenko, "Brain age prediction based on EEG records," 2022, Accessed: May 16, 2024. [Online]. Available: https://er.ucu.edu.ua/handle/1/3157

[40] V. R. Mehta, H. R. Parmar, A. B. Khandwala, K. M. Kakkad, V. V. Vekaria, and J. V. Bhatt, "A study of postnatal assessment of gestational age of neonates by new Ballards and Parkins score," *Int J Contemp Pediatr*, vol. 9, no. 1, p. 32, Dec. 2021, doi: 10.18203/2349-3291.ijcp20214849.

[41] N. J. Stevenson, Tone Nordvik, Cathrine Nygaard Espeland, Vito Giordano, Sissel J Moltu, Pål G Larsson, Katrin Klebermaß-Schrehof, Tom Stiris, and Sampsa Vanhatalo, "Inter-site generalizability of EEG based age prediction algorithms in the preterm infant," *Physiol. Meas.*, vol. 44, no. 7, p. 07NT01, Jul. 2023, doi: 10.1088/1361-6579/ace755.

[42] Soonhyun Yook, Yizhan Miao, Claire Park, Hea Ree Park, Jinyoung Kim, Diane C. Lim, Eun Yeon Joo, and Hosung Kim, "Predicting brain age based on sleep EEG and DenseNet," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Nov. 2021, pp. 245–248. doi: 10.1109/EMBC46164.2021.9631064.

[43] P. Kaushik, A. Gupta, P. Roy, and D. Dogra, "EEG-based Age and Gender Prediction Using Deep BLSTM-LSTM Network Model," *IEEE Sensors Journal*, vol. PP, pp. 1–1, Dec. 2018, doi: 10.1109/JSEN.2018.2885582.

[44] M. Khayretdinova, A. Shovkun, V. Degtyarev, A. Kiryasov, P. Pshonkovskaya, and I. Zakharov, "Predicting age from resting-state scalp EEG signals with deep convolutional neural networks on TD-brain dataset," *Frontiers in Aging Neuroscience*, vol. 14, 2022, Accessed: Jan. 23, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2022.1019869

[45] A. Ansari, K.Pillay, E.Arasteh, A.Dereymaeker, G.Schmidt Mellado, K.Jansen, A.M.Winkler, G.Naulaers, A.Bhatt, S.Van Huffel, C.Hartley, Maarten De Vos, R.Slater, and L.Baxter, "Resting state electroencephalographic brain activity in neonates can predict age and is indicative of neurodevelopmental outcome," In Review, preprint, Oct. 2023. doi: 10.21203/rs.3.rs-3487918/v1.

[46] A. Ansari, K.Pillay, L.Baxter, E.Arasteh, A.Dereymaeker, G.Schmidt Mellado, K.Jansen, G.Naulaers, A.Bhatt, S.Van Huffel, C.Hartley, Maarten De Vos, and R.Slater, "Brain age as an estimator of neurodevelopmental outcome: A deep learning approach for neonatal cot-side monitoring," Neuroscience, preprint, Jan. 2023. doi: 10.1101/2023.01.24.525361.

[47] L. Karlsson, M.Tolvanen, N.M Scheinin, H.M.Uusitupa, R.Korja, E.Ekholm, J.J.Tuulari, M.Pajulo, M.Huotilainen, T.Paunio, and Hasse Karlsson, "Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain)," *International Journal of Epidemiology*, vol. 47, no. 1, pp. 15–16j, Feb. 2018, doi: 10.1093/ije/dyx173.

[48] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," *Sci Rep*, vol. 11, no. 1, p. 6936, Mar. 2021, doi: 10.1038/s41598-021-86432-7.

[49] N. Junsomboon and T. Phienthrakul, "Combining Over-Sampling and Under-Sampling Techniques for Imbalance Dataset," in *Proceedings of the 9th International Conference on Machine Learning and Computing*, in ICMLC '17. New York, NY, USA: Association for Computing Machinery, Feb. 2017, pp. 243–247. doi: 10.1145/3055635.3056643.

[50] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[51] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery EEG-based classification: A review," *Biomedical Signal Processing and Control*, vol. 63, p. 102172, Jan. 2021, doi: 10.1016/j.bspc.2020.102172.

[52] V. H. Phung and E. J. Rhee, "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets," *Applied Sciences*, vol. 9, no. 21, Art. no. 21, Jan. 2019, doi: 10.3390/app9214500.

[53] A. Taherkhani, G. Cosma, and T. M. McGinnity, "A Deep Convolutional Neural Network for Time Series Classification with Intermediate Targets," *SN COMPUT. SCI.*, vol. 4, no. 6, p. 832, Oct. 2023, doi: 10.1007/s42979-023-02159-4.

[54] L. Alzubaidi, J.Zhang, A.J. Humaidi, A.Al-Dujaili, Y.Duan, O.Al-Shamma, J. Santamaría, M.A. Fadhel, M.Al-Amidie, and L.Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[55] T. Shi, L. Ren, and W. Cui, "Feature recognition of motor imaging EEG signals based on deep learning," *Pers Ubiquit Comput*, vol. 23, no. 3, pp. 499–510, Jul. 2019, doi: 10.1007/s00779-019-01250-z.

[56] J. Wang, S. Cheng, J. Tian, and Y. Gao, "A 2D CNN-LSTM hybrid algorithm using time series segments of EEG data for motor imagery classification," *Biomedical Signal Processing and Control*, vol. 83, p. 104627, May 2023, doi: 10.1016/j.bspc.2023.104627.

[57] V. K. Sudha and D. Kumar, "Hybrid CNN and LSTM Network For Heart Disease Prediction," *SN COMPUT. SCI.*, vol. 4, no. 2, p. 172, Jan. 2023, doi: 10.1007/s42979-022-01598-9.

[58] A. Seraj, and M Mohammadi-Khanaposhtani "Chapter 5 - Cross-validation," in *Handbook of Hydroinformatics*, S. Eslamian and F. Eslamian, Eds., Elsevier, 2023, pp. 89–105. doi: 10.1016/B978-0-12-821285-1.00021-X.

[59] B. E. Jaeger, S. Schmid, C. U. Grosse, A. Gögelein, and F. Elischberger, "Infrared Thermal Imaging-Based Turbine Blade Crack Classification Using Deep Learning," *J Nondestruct Eval*, vol. 41, no. 4, p. 74, Oct. 2022, doi: 10.1007/s10921-022-00907-9.

[60] R. Moradi, R. Berangi, and B. Minaei, "A survey of regularization strategies for deep models," *Artif Intell Rev*, vol. 53, no. 6, pp. 3947–3986, Aug. 2020, doi: 10.1007/s10462-019-09784-7.

[61] M. Dialameh, A. Hamzeh, H. Rahmani, S. Dialameh, and H. J. Kwon, "DL-Reg: A deep learning regularization technique using linear regression," *Expert Systems with Applications*, vol. 247, p. 123182, Aug. 2024, doi: 10.1016/j.eswa.2024.123182.

[62] M. Bentoumi, M. Daoud, M. Benaouali, and A. Taleb Ahmed, "Improvement of emotion recognition from facial images using deep learning and early stopping cross validation," *Multimed Tools Appl*, vol. 81, no. 21, pp. 29887–29917, Sep. 2022, doi: 10.1007/s11042-022-12058-0.

[63] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural Networks*, vol. 11, no. 4, pp. 761–767, Jun. 1998, doi: 10.1016/S0893-6080(98)00010-0.

[64] L. Rice, E. Wong, and Z. Kolter, "Overfitting in adversarially robust deep learning," in *Proceedings of the 37th International Conference on Machine Learning*, PMLR, Nov. 2020, pp. 8093–8104. Accessed: May 21, 2024. [Online]. Available: https://proceedings.mlr.press/v119/rice20a.html

[65] F. Kamalov and H. H. Leung, "Deep learning regularization in imbalanced data," in *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, Nov. 2020, pp. 1–5. doi: 10.1109/CCCI49893.2020.9256674.

[66] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, 2021, doi: 10.7717/peerj-cs.623.