MXLPred: A Natural Language Processing Deep Learning Model for ICU Patients' Data

Daphne Samuel
Department of Computer Science and Engineering, CEG Campus, Anna University, India
daphnesam7@gmail.com

Mary Anita Rajam
Department of Computer Science and Engineering, CEG Campus, Anna University, India
anitav@annauniv.edu

Abstract: *Early estimation of a patient's mortality rate and length of stay is crucial for saving lives. With the launch of Electronic Health Records (EHRs) in the healthcare industry, many studies have been done on clinical decision-making with EHR. The publicly accessible dataset, the Medical Information Mart for Intensive Care IV (MIMIC-IV) database, is used in this study. We propose a prediction of mortality and hospital stay length Mortality Prediction Ensemble Model with XLNet and GRU (MXLPred) framework to address the three main challenges of accurate medical condition prediction tools and data visualisation. First, we have created an ensemble model from eXtreme Language Net (XLNet) and Gated Recurrent Unit (GRU) learning algorithms to predict the first 24-hour and 90-day mortality rates during a patient's Intensive Care Unit (ICU) stay. We have utilised label-wise FastText embedding to identify medical entities from patient discharge summary notes, which has improved the clinical embedding and mortality estimates. We have also fine-tuned the model for better prediction results. Second, the Extreme Gradient Boosting (XGBoost) model for regression is utilised to predict the length of ICU stay of patients admitted in the ICU. Third, BigQuery is used to visualise the patient's current and past medical history to aid in quick decision-making. We compare the above prediction estimates with other state-of-the-art algorithms. We demonstrate that the final ensemble MXLPred shows improved outcomes regarding better Area Under the Receiver Operating Characteristic curve (AUROC) and Mean Squared Error (MSE).*

Keywords: *Deep learning, ensemble model, natural language processing, MIMIC-IV, regression, big query.*

Received September 15, 2024; accepted May 14, 2025
https://doi.org/10.34028/iajit/22/4/12

1. Introduction

Over the past years, Electronic Health Records (EHR) systems have been intended to store primary data associated with each patient. They were used for particular purposes like clinical informatics applications such as multi-outcome prediction, hospital readmission prediction, organ failure prediction, etc., As shown in Figure 1, classical EHR data comprises demographic information about patients, health records, sickness diagnoses, medications, vital sign diagnoses, immunisation records, test results, and radiology reports. EHRs can be used efficiently for various medical informatics studies, including risk assessments, early disease diagnosis, and predictive modelling.

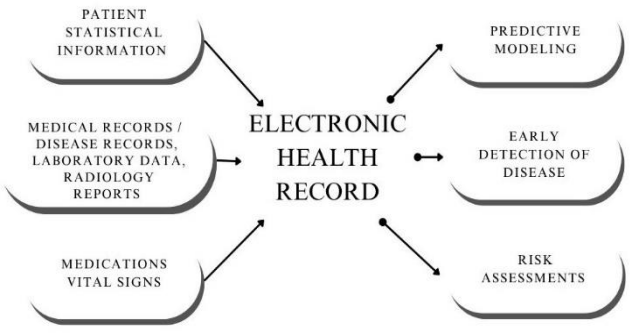


Figure 1. Electronic health record workflow.

The Intensive Care Unit (ICU) is a specialised medical facility that primarily aids patients in recovering from illnesses that pose a risk to their lives.

Medical examiners must continuously monitor patients in the ICU. Therefore, an early and accurate medical condition prediction tool would be an excellent mechanism to help treat patients just in time for their ailments. Here, deep learning techniques have been effectively used in learning features without human instruction, enabling the automatic detection of latent data associations that would otherwise go unnoticed.

The ability to reliably predict patient survival rates during and after an ICU stay is vital for intensive care decision-making and requires the ability to accurately forecast patient survival rates during and after an ICU stay. Several scoring techniques have been established for survival analysis and death prediction based on a patient's pulse rate and lab results. Yet, applying these standardised assessments to describe the severity of sickness and investigate patient deterioration that leads to high mortality risk during ICU admission is difficult [35]. International Classification of Diseases (ICD) coding is the process through which medical professionals assign ICD diagnosis codes to patients to record information about their medical diagnoses and procedures.

In this study, we introduce an ensemble model that

uses XLNet and Gated Recurrent Unit (GRU) to analyse a patient's clinical notes and forecast the likelihood of mortality for patients in the ICU. We use the Medical Information Mart for Intensive Care IV (MIMIC-IV) and MIMIC-IV discharge summary notes database, a restricted-access medical research repository data available on PhysioNet and managed by the Massachusetts Institute of Technology Laboratory (MIT Laboratory) for Computational Physiology[1]. Since it is of clinical significance, the MIMIC-IV database is used to create the patient cohort for our study, which consists of ICU patients older than 18 who have been in the ICU for at least one day for mortality prediction and length of stay (LoS) as a case study. The patient's ICU stay has static (e.g., gender, age) and time-series features.

The following is a summary of the significant contributions of this work:

- The cohort data is retrieved from the MIMIC-IV and MIMIC-IV discharge summary notes database to get information about the ICU patients.
- An enhanced clinical embedding technique is presented for the corpus of clinical discharge summary notes corpus, utilising label-wise FastText embedding.
- A novel ensemble model combining the XLNet and GRU models is developed to predict the ICU mortality rate.
- To predict the length of ICU stay, Extreme Gradient Boosting (XGBoost) regressor is utilised.
- To analyse and visualise data for descriptive and diagnostic purposes, BigQuery is used.

In the proposed Mortality Prediction Ensemble Model with XLNet and GRU (MXLPred) framework, we employ an ensemble architecture of XLNet and GRU that was not previously explored for ICU mortality prediction. Furthermore, we demonstrate the length of stay (LoS) prediction model as a regression task using an XGBoost regressor. In our work, we compare the model's effectiveness with state-of-the-art models. This paper is the first of its kind to use an XLNet and GRU ensemble approach in the MIMIC-IV database and BigQuery for data analysis and visualisation. The article is organised as follows for the remaining portion: Section 2 introduces the background facts on different clinical prediction models, various feature embeddings for mixed data, and studies concerning the MIMIC-IV data set. Section 3 highlights our motivations for cohort selection and novelty. Section 4 depicts the methodology in detail. Section 4.1 presents early patient identification of high mortality risks, which would allow for the prompt and appropriate delivery of medical care. The LoS predictions mentioned in section 4.2 could provide valuable information about patients. Section 4.3 explains the methods of creating visualisations with BigQuery in looker studio and exploring the data the query returns. Finally, sections 5 and 6 conclude with a presentation of the results and a discussion of the study.

2. Literature Survey

Accurately predicting a patient's mortality risk before, throughout, and following an ICU stay is crucial for assisting with decision-making in critical care. Various scoring methods, such as the Simpler Acute Physiology Score (SAPS) [19], Logistic Organ Dysfunction Score (LODS) [31], Acute Physiology and Chronic Health Evaluation III (APACHE III) [36] and Sequential Organ Failure Assessment (SOFA) [5] are utilised to support clinicians in ICU forecast prediction, to determine the mortality risk of patients. However, it is challenging to represent the entire level of medical severity with these standard scores and to examine patient deterioration connected with a higher chance of death during an ICU stay. Therefore, the severity level should be closely monitored throughout the ICU stay.

2.1. Feature Embedding

Electronic health records contain a wide range of patient health data. EHRs play a significant role in the ICD-coding process. For every patient interaction, clinicians create clinical notes, which are textual content archives. Beyond organised data like test results and medicines, clinical notes contain information that helps doctors provide better patient care. Because of their high dimensionality and lack of density, clinical notes have been underutilised in favour of more organised data. In Convolutional Attention for Multi-Label (CAML) classification, ICD-9 code prediction is treated as a multi-label binary text classification method [23]. The attention mechanism is applied to the most relevant code and passed to the output layer, with the likelihood of each code computed using a sigmoid transformation. Also, a regulariser is used to enhance code parameter similarity.

Semantic textual representations are learned using word embedding models based on their proximity to one another in a given context [3]. Some word embedding models are word2vec, FastText, Global Vectors for Word Representation (GloVe), and so on. Word embedding models suchas Word2Vec are 2-layer neural networks that learn the word representations in the text document in two different ways: asa constant bag-of-words and as a skip-gram [21]. Doc2Vec is a Word2Vec add-on that learns document-level embedding rather than word-level embedding. The Doc2Vec technique is then used to remember a low-rank vector from each patient's notes to receive the fixed-size vector of features. The clinical Named Entity Recognition (NER) algorithm [33] is used to keep track of the medically relevant keywords in the clinical notes prior to learning text-oriented representations. Comparable illness words, such as lymphadenitis and lymphoma, always have similar lexical structures. FastText can capture the additional medical data and the character-level n-gram representation of these similar subword associations.

2.2. Clinical Predictions

In the final hours of the patient's ICU stay, a switching state-space model links the patient's condition dynamics to the risk probabilities during a limited window (1-6 days) after discharge from the ICU. More precisely, by utilising a period in the sequential SAPS II outcome as a feature, an aggregate risk function evaluated the survival probability through an auto-regressive hidden Markov model [35].

Long-term temporal dependences have been captured in neural networks using Long Short-Term Memory (LSTM) [6], Gated Recurrent Unit (GRU), and XGBoost. These are some of the significant neural networks used for ICU predictions. Transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) [4] and XLNet have been mainly used when clinical notes are combined with structured EHR to find ICU predictions. These models have been vital in addressing complex collaborations among high-layered estimations. In order to learn text representation, BERT uses both sides (before targeting the token and after targeting the token) to anticipate the masked token. However, using this strategy does have certain drawbacks. This Masked Token (MASK) token is present during the training phase but not during the prediction phase. ClinicalBERT [10] has pre-trained BERT using clinical notes and has fine-tuned the network to predict hospital readmission. Clinical XLNet [11] analyses a patient's medical records to forecast the likelihood of prolonged mechanical ventilation and mortality using the permutation language modelling method suggested by XLNet on a corpus of discharge summaries to produce more accurate clinical embedding. Clinical text is used to fine-tune these models after they have been trained on medical data.

Khadanga *et al*. [17] utilised discharge summaries in addition to time-series data to enhance the prediction of ICU management benchmark tasks. Bardak *et al*. [1] used a modified approach from the one used by Mengqi Jin *et al*. [12] to extract time series features using the MIMIC-Extract [32] model and feed these features into GRU. They preprocessed clinical notes and used Med7 [18] to extract medical commodities. Then, Conventional Neural Network (CNN) was used to extract features from medical commodities representation. Finally, they concatenated time series and medical commodity features to predict mortality risk and LoS.

Medical Clinical Events Representation Learning-based FastText Time Long Short-Term Memory (MCPL-based FT-LSTM) [30], a technique for predicting clinical events based on learning medical concept representations, has used time-controlled LSTM to describe the fluctuating periods in EHRs, which captures better knowledge across long-term and short-term information and eliminates clinical data's significant dependency on timestamps.

To enhance ICU mortality prediction performance and deal with datasets with class imbalance, Pang et al. used the random down-sampling method for selection; however, this could lead to bias and possibly information loss [24]. MedFuse [8] uses an LSTM-based multimodal fusion approach for a combined clinical time series and chest X-ray image data in the MIMIC-IV and MIMIC-CXR. Previous studies demonstrated that using the SAPS II and SOFA scores as input variables, an ensemble learning algorithm might achieve a high degree of prediction [26, 27].

In recent years, generative language models such as Text-To-Text Transfer Transformer (T5) and BART have earned popularity due to their competitive execution in text creation and generative function modules. ClinicalT5 [20] presents a T5-based text-to-text transformer model that has been pre-trained on the clinical text and primarily inherits the objectives of T5 using unlabelled direct textual notes in the MIMIC-III database.

To summarise, although much research has been done to improve early prediction [25, 29], there are still many issues that need to be resolved, including how to correctly predict complex diseases and how to illustrate medical reports that are semantically comparable. Therefore, we propose the MXLPred framework, which ensembles XLNet and GRU, along with FastText [2] embedding. This model incorporates modelling methodologies from Auto-Encoder (AE) models Bidirectional Encoder Representations from Transformers (BERT) into Auto-Regressive (AR) models while avoiding the pitfalls of AE models. In contrast to traditional AR models, which use a predetermined forward or reverse factorisation order, XLNet maximises the anticipated log-likelihood of a series of overall potential factorisation order permutations. The permutation operation helps to learn bidirectional contextual information. By utilising the product rule to factorise the joint probability of the estimated tokens, the generalised AR language model removes the need for the assumption of independence in BERT. In addition, XLNet adds TransformerXL's segment recurrence mechanism and relative encoding schema to pretraining, which makes longer text sequences work better.

Here, FastText is utilised to learn the embedded expressions of medical notes, followed by the permutation language modelling techniques in XLNet and the time-efficient GRU model.

3. Cohort Selection

The statistics of the MIMIC-IV dataset are described in Table 1 The MIMIC-IV [13] and MIMIC-IV discharge summary notes [14] database are subjected to selecting a subset of data records to create our patient cohort, excluding irrelevant/redundant features. Our cohort consists of individuals, at least 18 years old, who have

been in the ICU for at least one day, averaging more than six hours per day. Those patients who are neither organ donors nor patients who have been moved from another hospital are added to our cohort. We also eliminate those with illnesses like neuromuscular disease, malignant tumours, severe burns, etc., that always necessitate protracted ICU stays to reduce ambiguity. Each ICU stay record has static (e.g., gender, age) and time-series features. This selected cohort is similar to the one used in Clinical XLNet [11], which uses the MIMIC-III database. The cohort selection criteria are detailed in Figure 2.

Table 1. Statistics of MIMIC-IV ICU dataset.

Number of stays	73,181
Unique patients	50,920
Hospital length of stay, (mean, SD)	11; 13.4
In-hospital mortality, (n,\%)	8,519; 11.6%

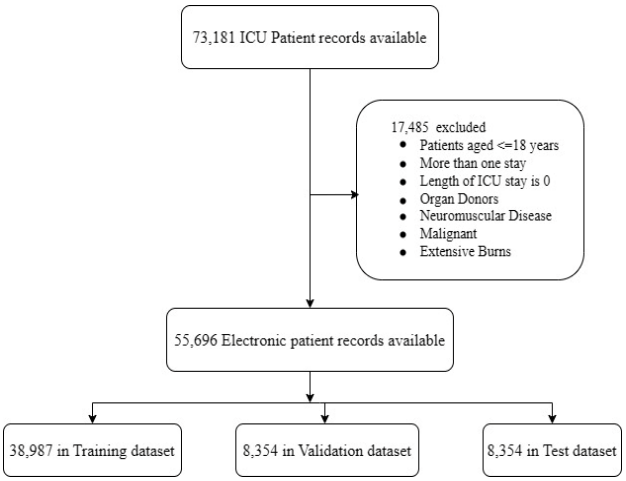


Figure 2. Cohort selection criteria.

A total of 73,181 unique ICU patient stays details exist in the MIMIC-IV database. From these, 17,485 ICU stays were excluded because they did not comply with the specifications in Figure 2. Finally, 55,696 are included in the cohort for analysis after the elimination of those with illnesses like neuromuscular disease, malignant tumours, serious burns, etc., that always necessitate prolonged ICU to reduce ambiguity.

From the tables in the MIMIC-IV database, the cohort data was extracted and summarised in Figure 3. The MIMIC-IV relational database comprises 35 individual tables, further organised into four distinct modules representing the core,hospital, ICU and derived tables. Age, marital status, departmental transfers, and admission details (including hospital) are all stored in the core module. Information such as laboratory results, microbiology, prescription administration, billable diagnoses/procedures, and physician orders are all accessible through the 'hosp' module, which gathers informationfrom the hospital-wide EHR. Accurate data, such as machine recordings and procedures, are stored in the ICU module after an ICU visit. The MIMIC code repository [22] for MIMIC-IV, which contains SQL codes to deliver clinically significant subgroup information like vasopressor medication, ventilation settings and blood gases, is used to retrieve the derived tables. We extract data using the selected cohort from the following tables: admissions, patients, and icu_stay. Also, we map the complete procedure_events tables, which hold routinely collected highly volatile data. ICD diagnosis codes are stored in the diagnosis_icd table and transferred to the cohort selection schema. Derived table adaptions are necessary to retrieve the hourly details of patients from ICU stay_hourly and patient routines from the vital sign table. We concatenate the discharge summary text field from the discharge table to form the final entire cohort.

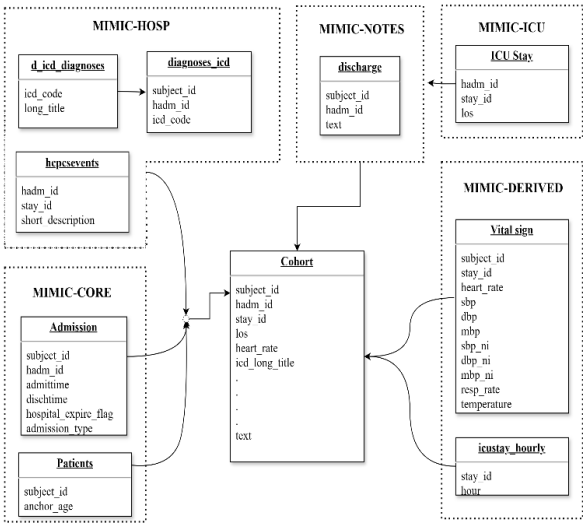


Figure 3. Data extraction from MIMIC-IV tables.

Outliers and imputation: outliers can significantly skew the results and disproportionately impact the performance of machine learning models. We utilise the MIMIC-Extract [32] variable-range table to create the outlier ranges and assign values to be imputed for each vital sign. This table helps to identify and manage outliers effectively by providing specific thresholds for each vital sign.

Table 2 details the accepted ranges for each vital sign so that outliers can be effectively identified and managed. For instance, heart rate values below 0 and above 350 are considered outliers, while the valid range is 0 to 350. An imputed value of 86 is used for missing data. By adhering to these ranges, we ensure the data used for model trainingand analysis remains realistic and clinically relevant.

Table 2. Vital signs variable ranges [32].

Vital signs	Outlier low	Valid low	Impute	Valid high	Outlier high
Glascow coma scale total	3	3	11	15	15
Heart rate	0	0	86	350	390
Systolic blood pressure	0	0	118	375	375
Diastolic blood pressure	0	0	59	375	375
Temperature	14.2	26	37	45	47
Respiratory rate	0	0	19	300	330
Weight	0	0	81.8	250	250
Haemoglobin	0	0	10.2	25	30

By defining these ranges, we can effectively filter out improbable or erroneous values, impute missing data with clinically relevant values, and maintain the integrity of the dataset. This process is crucial for enhancing the reliability and performance of our machine learning model, thereby ensuring extreme or inaccurate values do not excessively influence them.

4. Methodology

The MXLPred framework used to accomplish this paper's goal is illustrated in Figure 4. The first module collects the MIMIC-IV discharge summary notes based on cohort selection, which has been detailed in section 3. Then, it performs the ensemble XLNet, GRU and FastText embedding to identify individuals with high mortality risk. The second module utilises the cohort selection dataset and predicts the length of ICU stay. The third module uses the entire MIMIC-IV database to explore the insights into patients' mechanical ventilators and suspicion of infection with the antibiotics administered data that the SQL query returns. This section is organised as follows: Section 4.1 presents the details of the ensemble XLNet and GRU units used for the ICU mortality prediction model. Section 4.2 describes the model used for the prediction of the LoS task. Section 4.3 explains different data visualisation and analysis techniques for time series data extracted from the dataset using BigQuery.

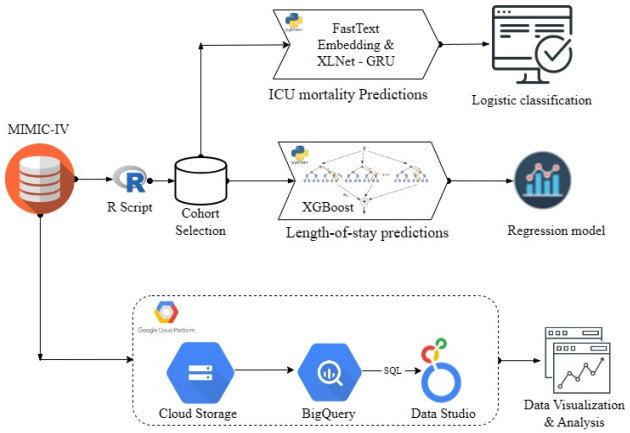


Figure 4. MXLPred workflow block diagram.

4.1. ICU Mortality Predictions

The architecture of the MXLPred model proposed for ICU mortality predictions is shown in Figure 5, which illustrates a sample workflow of information extraction for a patient who is diagnosed with an abscess of the anal and rectal regions. The proposed ensemble model has four layers, as described in Algorithm (1). In the first layer, label-wise is used to obtain representations corresponding to the FastText embeddings of each patient record. In the second layer of XLNet, all the input tokens are transformed into latent feature representations. The third layer is a GRU layer, which creates weight vectors with specified labels, each depicting the entire input text. The final layer for fine-tuning the model comprises target-specific binary classifiers on top of the matching target-specific vectors.

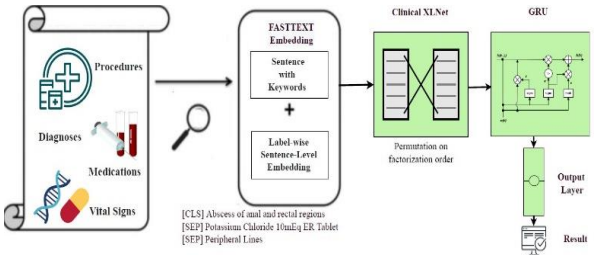


Figure 5. The architecture of the proposed XLNet-GRU ensemble model.

Algorithm 1: Predict First 24-Hour and 90-Day Mortality

1: Initialization ⟸ Embedding dimention, Learning rate, Batchsize
2: ***function*** *SEQUENTAL MODELING*
3: Label-wise Fast Text embedding
4: Train the XLNet Model (hidden layer size, batch size)
5: GRU (hidden layer size, batch size)
6: Drop out layer (0.25)
7: ReLU
8: Predector Neural Network Layer (hidden layer size, batch size)
9: end function

The final layer for fine-tuning the model comprises target-specific binaryclassifiers on top of the matching target-specific vectors.

4.1.1. Embedding Layer

The FastText model achieves accurate word vectors, considering the word's underlying structure and syntactic information. It can derive vectors for Out-Of-Vocabulary (OOV) words by aggregating vectors from its component character-level n-grams, given that at least one of the character n-grams existed in the training data.

As a result,we use the FastText model, which better suits the represented medical events. Each patient is represented by P, and each patient is connected to a chronologically ordered series of notes N_i. To anticipate the words that will appear in the context, FastText uses the skip-gram model. Consider the word vocabulary size to be F, when a word's index serves as a reference j∈1, …, F the objective is to understand each word's f vectorised equivalent. Words f_1, …, and f_T are used to denote the training corpus. The logarithmic likelihood function given below is finally optimised:

$$L = \sum_{t=1}^{T} \sum_{c \in C_t} log\ p(f_c|f_t) \quad (1)$$

Let the size of the clinical notion vector be T, the size of the sliding window be c, the centre theme word be f_t the group of words that around f_t be C_t and the contextual theme word for f_t be f_c.

The softmax function may modify $p(f_c|f_t)$ in the following expression, where s stands for the word

similarity function.

The representation of each word *f* is a bag containing character *n*-grams. This prevents the internal arrangement of words from being ignored. However, there are similar spelling words that contain numerous identical subword components. Yet, these associated subword connections and related clinical conceptual information may be effectively captured by this subword formulation. Subvector expression improves predictions' performance and further improves training precision.

$$p(f_c|f_t) = \frac{e^{s(f_t,f_c)}}{\sum_{i=1}^{W} e^{s(f_t,f_i)}} \quad (2)$$

Consider the subword corpus length to be B. The n-gram occurring in *f* is indicated by $B_f \subset \{1, \ldots, B\}$, and J_b is used to vectorise the subword b in each n-gram. The vector form of the word f_c is v_{f_c}. As a result, the scoring function is acquired in Equation (3). Following that, the vector sum of a word's *n*-grams serves as a representation of the word.

$$s(f_c, f_t) = \sum_{b \in B_{f_t}} J_b^{\intercal} v_{f_c} \quad (3)$$

Subsequently, we apply the XLNet model to make further predictions after feeding the resulting numerical vector into it.

4.1.2. XLNet Layer

A variant of the transformer-xl model is XLNet. The generalised auto-regressive language model XLNet learns an unsupervised representation of text sequences. This model avoids the drawbacks of AE while incorporating modelling strategies from AE models BERT into AR models.

A permutation is done in factorisation orders since the sequence of original words corresponds to the positional encoding, and they rely on proper attention masks in Transformers. Two-stream self-attention allows the prediction to be aware of the target position. One of the few models without a restriction on sequence length is XLNet.

In XLNet downstream jobs, a Classification Token (CLS) classification token is added at the start of every series of words. Permutation Language modelling initially creates a list of order-factorised sequences from the input sequence. After that, language modelling is used to anticipate the subsequent words based on the preceding ones. Therefore, given a patient's associated temporally ordered sequence of notes, we acquire a single piece of note representation as output.

$$\hat{\theta} = argmax_{\theta} \left[\mathbb{E}_{k \sim K}\left[\sum_{t=1}^{T} log\ P_{\theta}[(z_{k_t}|z_{k_{<t}})]\right]\right] \quad (4)$$

XLNet uses the principle of Equation (4) as a goal and optimises the model parameters θ to maximise the probability of tokens Z_{k_t} in an order of a sequence *T* for all possible permutations of K_T.

An AR model determines the likelihood of preceding tokens $Z_{k<t}$, where k_t is the element at position *t* in a permutation *k* of the token index and $k_{<t}$ are the elements preceding to actually *t* in the permutation. As the predictive value is a product function of the likelihood for every component in the sequence, the sum of the log-likelihood suggests that the model is adequately auto-regressive for each permutation. As demonstrated by the expectation across all permutations in *K*, the model has been trained to be equally capable of computing probability for each token, given any context.

4.1.3. GRU Layer

The sequential modelling layer takes advantage of the temporal information found among the notes. GRU is fast and accurate because of its simple structure. To address the vanishing gradient issue that traditional Recurrent Neural Networks (RNNs) encounter, the reset gate and update gate are incorporated. GRU has fewer training parameters than other RNNs, which results in less memory usage, faster execution, and quicker training.

In the GRU unit, g_t symbolises input at time step *t*, $x(t\text{-}1)$ symbolises the hidden state from the previous time step, r_t symbolises the reset gate vector, k_t symbolises the update gate vector, x_t symbolises the output vector, n_t symbolises candidate activation vector, *F* symbolises the parameter weight, the logistic sigmoid function σ of range (0, 1), and b is the bias.

$$r_t = \sigma\left(F_{ir} g_t + b_{ir} + F_{xr} x_{(t-1)} + b_{xr}\right) \quad (5)$$

$$k_t = \sigma\left(F_{ik} g_t + b_{ik} + F_{xk} x_{(t-1)} + b_{xk}\right) \quad (6)$$

$$n_t = \tanh\left(F_{in} g_t + b_{in} + r_t * \left(F_{xn} x_{(t-1)} + b_{xn}\right)\right) \quad (7)$$

$$x_t = (1 - k_t) * n_t + k_t * x_{(t-1)} \quad (8)$$

4.1.4. Output Layer

The output of the XLNet-GRU is used for classification in the predictor neural network, which ultimately produces a probability that assesses the likelihood of the downstream target variables. To reduce algorithmic errors, neural networks employ optimising techniques. We apply the binary cross-entropy loss function in Equation (9), to calculate this loss for *N* samples with y_t true label, and $!\,y_t'$ predicted label. It is used to express how well the model is working.

$$L(y_t, y_t') = \frac{1}{N} \sum_{1 \le t \le N} (y_t! \times !\ln y_t'! + !\,(1! - !\,y_t)! \times !\ln(1! - !\,y_t')) \quad (9)$$

4.2. Length of Stay (LoS) Predictions

The supervised regression technique known as XGBoost efficiently implements the gradient boosting approach. Gradient boosting refers to a class of ensembles constructed from decision tree machine

learning algorithms. Gupta *et al.*'s [7] MIMIC-IV pipeline evaluates a classification model for LoS>3, whereas we handle the LoS prediction as a regression task using the XGBoost regressor. One tree would be added to the ensemble at a time and fitted to the correct predictions the earlier models made incorrectly. The two key benefits of XGBoost are model performance and execution speed. Further, by analysing the deep learning models with XGBoost on many datasets, current research has demonstrated that deep models are not required for information that is tabulated. However, an ensemble of XGBoost along with deep learning modelachieves maximum performance [28].

Mean Squared Error (MSE), mean Pinball Loss (PbL), and R^2 score are the performance indicators used for n number of samples, y_i observed LoS, and $\hat{y}_i$ predicted LoS. These are some key metrics of XGBoost models, and each plays an important role.

Mean Squared Error: MSE in Equation (10) is the mean of the squares of the errors $(y_i - \hat{y}_i)^2$ as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (10)$$

Mean Pinball Loss: the efficiency of quantile regression models is assessed using the PbL function in Equation (11), where α is the slope of the pinball loss, and its default value is 0.5.

$$\text{Pinball}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1} \alpha \max(y_i - \hat{y}_i, 0) + (1 - \alpha)\max(\hat{y}_i - y_i, 0) \quad (11)$$

R^2 score: a statistical indicator of how well a regression model fits the data is R-squared in Equation (12), where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. The better the model is fitted, the closer the r-square value is to 1.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (12)$$

4.3. Visualisation with BigQuery

BigQuery is a completely automated data storage with built-in tools for managing and analysing data, including business intelligence, machine learning, and geographic analysis. SQL queries may be executed with BigQuery's serverless design, which eliminates the need for infrastructure management. BigQuery integrates powerful analytical capabilities with a cloud-hosted data warehouse.

In BigQuery, the SQL query is the central analytical unit. The queries are executed in the Google Cloud Platform (GCP) console integrated with BigQuery. Data is loaded into BigQuery for analysis from cloud storage. The MIMIC-IV database is accessed from the data hosted on PhysioNet cloud storage.

We use BigQuery for descriptive and diagnostic analytics. For example, based on the vital signs data [15] for a particular patient ID, the time series chart is depicted in Figure 6. We can infer from Figure 6 that on 7th March, the Systolic Blood Pressure (SBP) decreased and the heart rate increased. Eventually, the patient was moved to a ventilator for invasive ventilation therapy. This patient was diagnosed with a positive infection for a respiratory culture on 5th and 17th. The medications given were antibiotics.

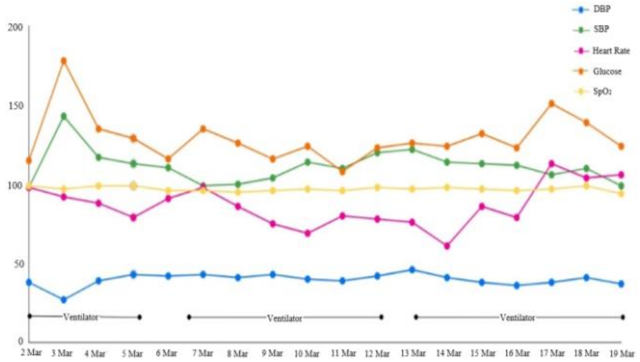


Figure 6. BigQuery vital sign: time series analysis.

5. Results and Discussion

In this section, we evaluate and discuss the results of the three goals of our study that may be useful in assessing the prediction of risk in patients. The evaluation metrics are described in detail. Then, the experimental setup for network training is described to reduce the risk of patient mortality. The outcomes of this model would be discrete, depending on whether the patient has survived or has died. Also, model performance analysis and the effects of hyperparameter tuning are discussed. Further, we forecast the length of ICU stay tasks and compare their prediction estimates with other state-of-the-art algorithms. Finally, we report the visualisation and analysis with BigQuery.

5.1. Evaluation Metrics

After training machine learning classifiers with sufficient training samples, the next step is to present the classifier with unfamiliar or test samples and determine whether or not the samples are correctly categorised. The outcome of this classification must be assessed and quantified in some way.

The basis of accuracy, precision, recall, F1-Score, and AUROC comes from the concepts of true positive, true negative, false positive, and false negative.

- True Positive (TP): survived patients' instances that were correctly identified as survived.
- False Positive (FP): deceased patients' instances that were misclassified as positive.
- True Negative (TN): instances of deceased patients that were correctly identified as negative.
- False Negative (FN): instances of survived patients that were misclassified as deceased.

- Confusion Matrix: to compare the actual target instances with those the machine learning model predicted, one uses an NxN matrix called a confusion

matrix.

- Precision: precision can be defined as the proportion of correctly predicted positive results. The precision of a model is 1.0 if it generates no false positives. The formula is as follows:

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{13}$$

- Recall: the capacity to recognise each relevant value in the data collection is known as recall.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

- Accuracy: accuracy describes the number of correct and overall predictions.

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{Number of test samples}} \tag{15}$$

- F1-Score: the F1 score is depicted as the harmonic mean of recall and accuracy. It is the measure used in statistical factors to rate the model's performance. Since the F1 score incorporates precision and recall into a single metric, having one performance metric rather than multiple is much more convenient.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{16}$$

- AUROC: the region beneath the ROC curve is known as the AUROC for a certain curve. The optimal AUROC is 1, whereas the lowest one is 0.5. The trade-off between TP and FP at various decision thresholds between 1 and 0 is displayed by the AUROC curve. When dealing with unbalanced data, this statistic provides greater insight.

5.2. ICU Mortality Predictions

The pre-training process was conducted on a workstation with an Intel® XeonTM processor, an NVIDIA Quadro P4000 graphics card, and 64 GB of RAM.

5.2.1. Training Details

Each patient's diagnostic records are retrieved and entered chronologically into the FastText model, where we configure the FastText embedding with a window size of 5, a character-level subword of 3, and select a 300-dimensional vector representation for each patient. For downstream tasks at the XLNet layer, we employ a 32-batch size with a learning rate of 1e-5 for four epochs, implementing early stopping for efficient training. Following this, we incorporate a GRU module with a batch size of 128 and a learning rate of 2e-5 to fine-tune the model. The Adam optimiser is applied with an epsilon of 1e-6 and a learning rate 2e-5.

5.2.2. Model Comparison

The patients' data is randomly split into 70% training and 30% testing set. Fivefold cross-validation is used to find and evaluate the best hyperparameters. Seven state-of-the-art machine learning models are compared with our proposed MXLPred ensemble XLNet-GRU to predict ICU mortality with AUROC.

Table 3 shows that the XLNet-GRU ensemble has a better AUROC and outperforms the baseline scoring method for the different tasks. Our MXLPred model achieves better performance than state-of-the-art results, and the combined learning method aids in improving performance for OOV words that are used infrequently. Figure 7 compares the AUROC results for the various models for the 90-day mortality prediction task.

Table 3. Performance in the prediction task for mortality risk.

Model	90-Day mortality		24-Hour mortality	
	AUROC	AUPRC	AUROC	AUPRC
Random forest	0.745	0.52	0.673	0.49
BERT [4]	0.784	0.63	0.733	0.56
Med2Vec [34]	0.812	0.65	0.705	0.54
Clinical BERT [10]	0.833	0.69	0.755	0.57
Clinical XLNet [11]	0.851	0.72	0.792	0.64
Clinical NotesICU [17]	0.852	0.72	0.790	0.64
Convolution MedicalNER [1]	0.865	0.71	0.804	0.66
Proposed ensemble	0.876	0.74	0.843	0.69

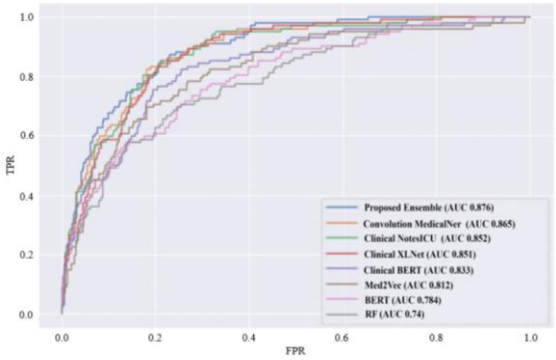


Figure 7. AUROC on the 90-day mortality prediction task for classification models.

In Figure 8, the confusion matrix is shown for the 24-hour mortality prediction model. The model correctly predicted that 2.90% of positive instances were positive, while 4.42% of positive instances were misclassified as negative by the model.

Figure 8. Confusion matrix for 24-hour mortality.

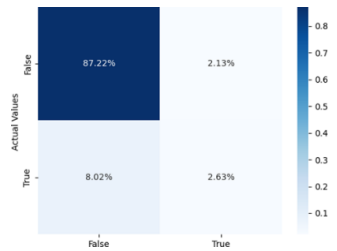
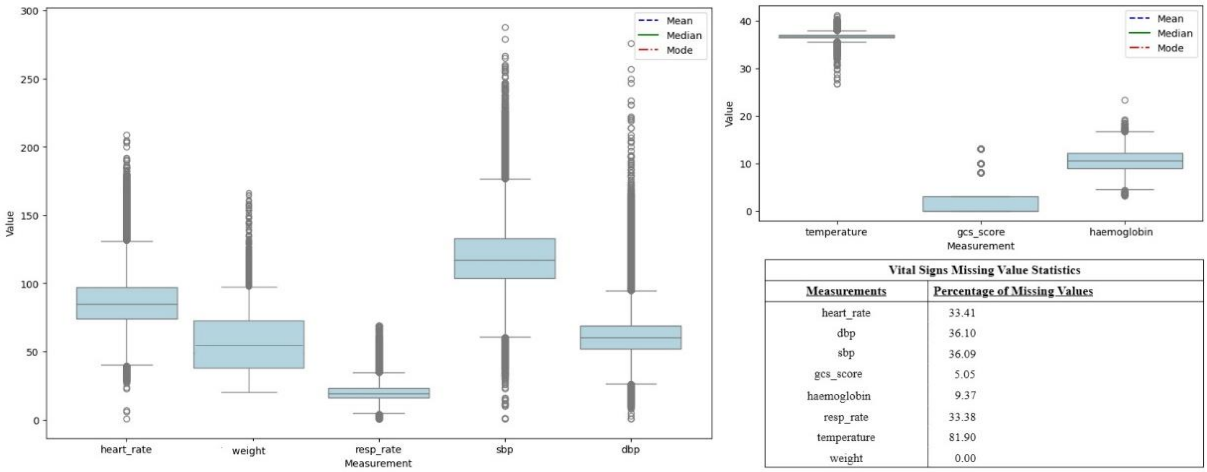


Figure 9. Confusion matrix for 90-day mortality.

In the 90-day mortality predictions model, 2.63% of positive instances were correctly predicted as positive by the model, while the 8.02% of positive instances were misclassified as negative by the model. The confusion matrix is shown in Figure 9.

5.3. Length of Stay Predictions

Figure 10 illustrates the vital sign distribution as a box plot, highlighting the mean, median, mode, and outliers. This plot provides a view of the central tendency and variability of the most essential features and highlights the existence of extreme values.



Vital Signs Missing Value Statistics

Measurements	Percentage of Missing Values
heart_rate	33.41
dbp	36.10
sbp	36.09
gcs_score	5.05
haemoglobin	9.37
resp_rate	33.38
temperature	81.90
weight	0.00

Figure 10. Box plot of vital signs showing central tendencies and outliers.

The median, represented by the line insidethe box, denotes the 50th percentile or the middle value of the dataset. The mean, typically marked with a distinct lineor point, provides the average value. At the same time, the mode, often indicated as a separate point, shows the most frequently occurring value in the dataset. The Interquartilerange (IQR), represented by the box, highlights the spread of the central values and helps understand the data's dispersion. Furthermore, Figure 10 shows the percentage of missing values for patients' vital signs, which provides additional insight into data quality.

Table 4 shows the performance measured for all modelsin the LoS prediction task from the MIMIC-IV database.

Table 4. Performance of length of stay prediction.

Model	MSE	PbL	R^2
XGBoost regressor	4.943	0.461	0.98
DNN [16]	5.112	0.501	0.98
1D-CNN [16]	5.267	0.514	0.97
Random forest	5.463	0.523	0.96
Gradient boosting regressor	12.254	0.612	0.94
SVM	13.243	0.546	0.93
KNN regressor	13.382	0.431	0.94
Decision tree	13.462	0.442	0.94
Bayesian regressor	13.882	0.437	0.94

The XGBoost regression model exhibits the best performance regarding MSE, PbL and the R2 score.

In Figure 11, the rankings of selected features for predicting LoS are presented. The model prioritised age as the most important feature. This finding aligns with the observations by Hol *et al*. [9], who noted a correlation between increasing age and the likelihood of complications, extended hospital stays in the ICU, and higher mortality rates.

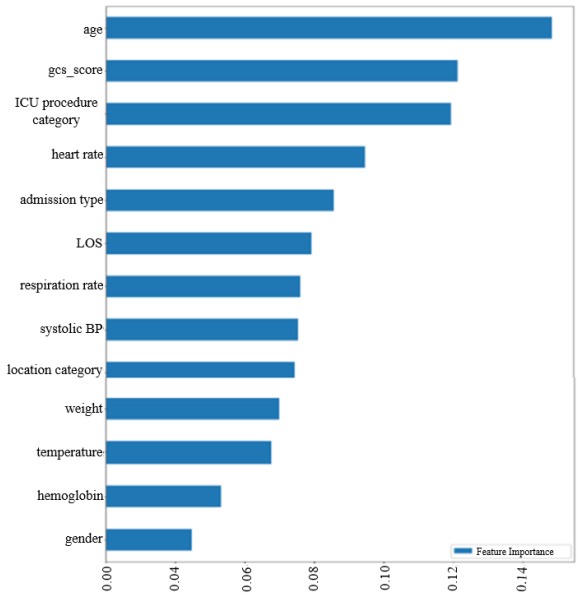


Figure 11. Important feature selection for the length of stay predictions.

5.4. Visualisation with BigQuery

The SQL dialects are executed with BigQuery. We access the MIMIC-IV hosted data from physionet using BigQuery on GCP. We explore BigQuery data using Looker Studio to know the impact of a patient's age on how well they respond to medications with prolonged mechanical ventilation. Looker Studio is an enterprise intellect platform that lets users build and consume data visualisations, dashboards, and reports. With Looker Studio, we explore and visualise the query results and tabular data in a Looker Studio report. Sharing our insights with others is also easy. These data are

aggregated and derived to analyse the correlation between LoS and ventilator duration, as discussed in the following section. In addition, a brief analysis of the various antibiotics used in ICU patients suspected of infection.

5.4.1. Inference from Ventilator Data

According to Figure 12, the effect of ageing on ventilator patients and their correlation with LoS are understood. Figure 13 is obtained from the derived table ventilation; we infer that supplemental oxygen is the most frequent administration prescribed by health care providers for patients on ventilators. Supplemental oxygen is frequently given to severely ill patients on mechanical ventilators to treat hypoxemia. Supplemental oxygen, invasive ventilation, non-invasive ventilation, tracheostomy, and High-Flow Nasal Cannula (HFNC) oxygen therapy are the different treatments given to patients on ventilators. Additionally, we infer that a higher number of ventilator patients are in the age group 58-77.

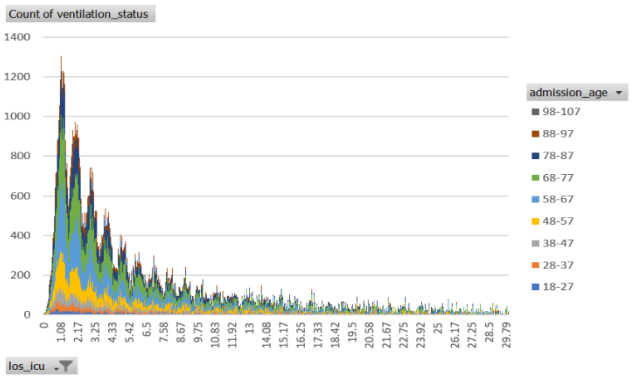


Figure 12. LoS for the different age groups in mechanical ventilator.

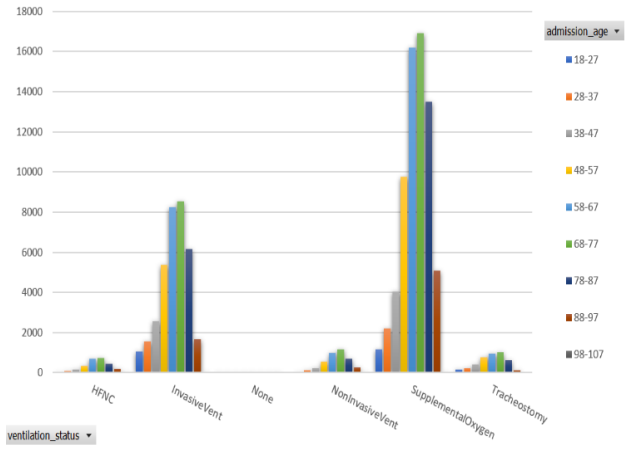


Figure 13. Comparative analysis between ventilators types and age groups.

A stacked column chart was created to demonstrate the drugs most frequently used during a patient's stay in the ICU to highlight the impact of antibiotics associated with suspicion of infection. Vancomycin is the antibiotic most preferred by clinicians for the suspicion of infection, which can be inferred from Figure 14.

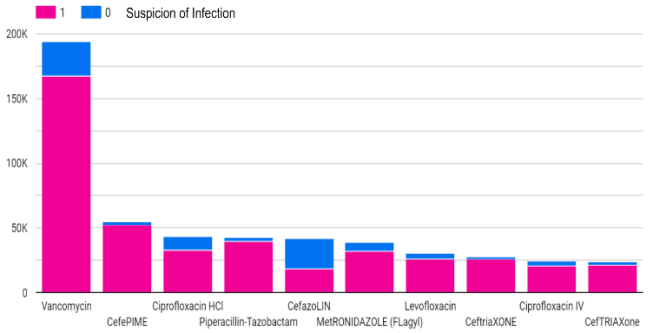


Figure 14. Stacked column chart shows the ranking of antibiotics used for suspicion of infection.

The other most commonly used antibiotics are Cefepime, Ciprofloxacin HCL, Piperacillin-Tazobactam, Cefazolin, Metronidazole (Flagyl), Levofloxacin, Ceftriaxone, Ciprofloxacin IV, and so on.

Supplementary data shows further analysis of the charts for antibiotics involved in various stages of the investigation and different organ failure outcome scores.

6. Conclusions and Future Work

We were motivated to work on mortality prediction and length of stay for ICU patients. This critical clinical issue necessitates improved methods for anticipating complications, saving lives, and optimising resource allocation in healthcare decision-making. Finding the best hyperparameters for deep learning models remains a challenge. Exhaustive search methods in the search space are tedious and time-consuming.

We present enhanced clinical embedding using FastText embedding on the corpus of clinical discharge summary notes. The proposed MXLPred Mortality Prediction model using the ensemble XLNet-GRU architecture achieved significant improvement in AUROC and accuracy compared to state-of-the-art techniques for ICU mortality risk prediction on the MIMIC-IV dataset.

This ensemble model effectively reduces generalisation error without increasing variance, a key challenge in deep learning models. The MXLPred LoS prediction using XGBoost regression predicts the LoS quite accurately compared to the other models.

In essence, the MXLPred classification and regression models performed best in predicting mortality and LoS, with better AUROC and MSE scores. Further, based on the data visualisation and analytics results on BigQuery, we infer that supplemental oxygen and Vancomycin antibiotics are the most frequent medications prescribed by healthcare providers for patients on ventilators and with suspicion of infection, respectively.

There are many ways to improve this research. Future work can be carried out in the following directions: During a pandemic like COVID-19, hospitals can face challenges with data availability. To address this, Federated Learning (FL), a distributed AI technique,

could enable model training without sharing raw data. This would involve coordination with various clients (such as hospitals). Also, our study did not explore functionalities like question-answering and other associated tasks for clinical documents.

Acknowledgment

This work is supported by the Anna Centenary Research Fellowship [CFR/ACRF/21244191591/AR1, Dated: 19-02-2021].

Data Availability Statement

The data underlying the results presented in the work are available from the Medical Information Mart for Intensive Care (MIMIC-IV), a large, single-centre database comprising information relating to patients admitted to critical care units at a large tertiary care hospital. More information about MIMIC-IV can be found on their website. To access these data, interested researchers must first complete the CITI 'Data or Specimens Only Research' course and then apply for credentialed access through PhysioNet.

References

[1] Bardak B. and Tan M., "Improving Clinical Outcome Predictions Using Convolution over Medical Entities with Multimodal Learning," *Artificial Intelligence in Medicine*, vol. 117, pp. 102112, 2021. https://doi.org/10.1016/j.artmed.2021.102112

[2] Bojanowski P., Grave E., Joulin A., and Mikolov T., "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, 2017. https://doi.org/10.1162/tacl_a_00051

[3] Bourahouat G., Abourezq M., and Daoudi N., "Word Embedding as a Semantic Feature Extraction Technique in Arabic Natural Language Processing: An Overview," *The International Arab Journal of Information Technology*, vol. 21, no. 02, pp. 313-325, 2024. https://doi.org/10.34028/iajit/21/2/13

[4] Devlin J., Chang M., and Lee K., "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *arXiv Preprint*, vol. arXiv:1810.04805v2, 2019. https://arxiv.org/abs/1810.04805v2

[5] Elias A., Agbarieh R., Saliba W., Khoury J., Bahouth F., Nashashibi J., and Azzam Z., "SOFA Score and Short-Term Mortality in Acute Decompensated Heart Failure," *Scientific Reports*, vol. 10, no. 1, pp. 1-10, 2020. https://doi.org/10.1038/s41598-020-77967-2

[6] Ge X., Huh J., Park Y., Lee J., Kim Y., and Turchin A., "An Interpretable ICU Mortality Prediction Model Based on Logistic Regression and Recurrent Neural Networks with LSTM Units," *AMIA Annual Symposium Proceedings AMIA Symposium*, vol. 2018, pp. 460-9, 2018. https://pmc.ncbi.nlm.nih.gov/articles/PMC6371274/pdf/2974987.pdf

[7] Gupta M., Gallamoza B., Cutrona N., Dhakal P., Poulain R., and Beheshti R., "An Extensive Data Processing Pipeline for MIMIC-IV," *in Proceedings of the 2nd Machine Learning for Health Symposium Conference*, New Orleans, pp. 311-325, 2022. https://proceedings.mlr.press/v193/gupta22a.html

[8] Hayat N., Geras J., and Shamout F., "MedFuse: Multimodal Fusion with Clinical Time-Series Data and Chest X-Ray Images," *arXiv Preprint*, vol. 182, pp. 1-25, 2022. https://arxiv.org/abs/2207.07027

[9] Hol L., Van Oosten P., Nijbroek S., Tsonas A., Botta M., Neto A., Paulus F., and Schultz M., "The Effect of Age on Ventilation Management and Clinical Outcomes in Critically Ill COVID-19 Patients-Insights from the PRoVENT-COVID Study," *Aging*, vol. 14, no. 3, pp. 1087-1109, 2022. DOI:10.18632/aging.203863

[10] Huang K., Altosaar J., and Ranganath R., "ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission," *arXiv Preprint*, vol. arXiv:1904.05342v3, pp. 1-9, 2020. http://arxiv.org/abs/1904.05342

[11] Huang K., Singh A., Chen S., Moseley E., Deng C., George N., Deng C., George N., and Lindvall C., "Clinical XLNet: Modeling Sequential Clinical Notes and Predicting Prolonged Mechanical Ventilation," *arXiv Preprint*, vol. arXiv:1912.11975v1, pp. 1-9, 2019. http://arxiv.org/abs/1912.11975

[12] Jin M., Bahadori M., Colak A., Bhatia P., Celikkaya B., Bhakta R., Senthivel S., Khalilia M., Navarro D., and Zhang B., "Improving Hospital Mortality Prediction with Medical Named Entities and Multimodal Learning," *arXiv Preprint*, vol. arXiv:1811.12276v2, pp. 1-8, 2018. https://doi.org/10.48550/arXiv.1811.12276

[13] Johnson A., Bulgarelli L., Shen L., Gayles A., Shammout A., Horng S., Pollard T., Hao S., Moody B., and Gow B., "MIMIC-IV, A Freely Accessible Electronic Health Record Dataset," *Scientific Data*, vol. 10, no. 1, pp. 1-9, 2023. https://doi.org/10.1038/s41597-022-01899-x

[14] Johnson A., Pollard T., Horng S., Celi L., Mark R., "MIMIC-IV-Note: Deidentified Free-Text Clinical Notes," *PhysioNet*, 2023. https://physionet.org/content/mimic-iv-note/2.2/

[15] Johnson A., Stone D., Celi L., and Pollard T., "The MIMIC Code Repository: Enabling Reproducibility in Critical Care Research," *Journal of the American Medical Informatics Association*, vol. 25, no. 1, pp. 32-39, 2018.

DOI:10.1093/jamia/ocx084

[16] Kaggle-MoA-2nd-Place-Solution/dnn-train.ipynb at Main baosenguo/Kaggle-MoA-2nd-Place-Solution GitHub, https://github.com/baosenguo/Kaggle-MoA-2nd-Place-Solution/blob/main/training/dnn-train.ipynb, Last Visited, 2024.

[17] Khadanga S., Aggarwal K., Joty S., and Srivastava J., "Using Clinical Notes with Time Series Data for ICU Management," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, pp. 6432-6437, 2019. DOI:10.18653/v1/D19-1678

[18] Kormilitzin A., Vaci N., Liu Q., and Nevado-Holgado A., "Med7: A Transferable Clinical Natural Language Processing Model for Electronic Health Records," *Artificial Intelligence in Medicine*, vol. 118, pp. 102086, 2021. https://doi.org/10.1016/j.artmed.2021.102086

[19] Lee H., Yoon S., Oh S., Shin J., Kim J., Jung C., and Ryu H., "Comparison of APACHE IV with APACHE II, SAPS 3, MELD, MELD-Na, and CTP Scores in Predicting Mortality After Liver Transplantation," *Scientific Reports*, vol. 7, no. 1, pp. 1-10, 2017. https://www.nature.com/articles/s41598-017-07797-2

[20] Lu Q., Dou D., and Nguyen T., "ClinicalT5: A Generative Language Model for Clinical Text," *in Proceedings of the Findings of the Association for Computational Linguistics*, Abu Dhabi, pp. 5436-5443, 2022. DOI:10.18653/v1/2022.findings-emnlp.398

[21] Mikolov T., Chen K., Corrado G., and Dean J., "Efficient Estimation of Word Representations in Vector Space," *arXiv Preprint*, vol. arXiv:1301.3781v3, pp. 1-12, 2013. https://arxiv.org/abs/1301.3781v3

[22] MIT-LCP, MIT-LCP/Mimic-Code: Mimic Code Repository: Code Shared by the Research Community for the Mimic Family of Databases, https://github.com/MIT-LCP/mimic-code, Last Visited, 2024.

[23] Mullenbach J., Wiegreffe S., Duke J., Sun J., and Eisenstein J., "Explainable Prediction of Medical Codes from Clinical Text," *in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, pp. 1101-1111, 2018. DOI:10.18653/v1/N18-1100

[24] Pang K., Li L., Ouyang W., Liu X., and Tang Y., "Establishment of ICU Mortality Risk Prediction Models with Machine Learning Algorithm Using MIMIC-IV Database," *Diagnostics*, vol. 12, pp. 1068, 2022. https://doi.org/10.3390/diagnostics12051068

[25] Parveen H., Rizvi S., and Shukla P., "Disease Risk Level Prediction Based on Knowledge Driven Optimized Deep Ensemble Framework," *Biomedical Signal Processing and Control*, vol. 79, pp. 103991, 2023. https://doi.org/10.1016/j.bspc.2022.103991

[26] Pirracchio R., Petersen M., Carone M., Rigon M., Chevret S., and Van der Laan M., "Mortality Prediction in Intensive Care Units with the Super ICU Learner Algorithm (SICULA): A Population-Based Study," *The Lancet Respiratory Medicine*, vol. 3, no. 1, pp. 42-52, 2015. DOI: 10.1016/S2213-2600(14)70239-5

[27] Purushotham S., Meng C., Che Z., and Liu Y., "Benchmarking Deep Learning Models on Large Healthcare Datasets," *Journal of Biomedical Informatics*, vol. 83, pp. 112-134, 2018. https://doi.org/10.1016/j.jbi.2018.04.007

[28] Shwartz-Ziv R. and Armon A., "Tabular Data: Deep Learning is Not All You Need," *Information Fusion*, vol. 81, pp. 84-90, 2021. https://doi.org/10.1016/j.inffus.2021.11.011

[29] Wang H., Wang C., Xu J., Yuan J., Liu G., and Zhang G., "Invasive Mechanical Ventilation Probability Estimation Using Machine Learning Methods Based on Non-Invasive Parameters," *Biomedical Signal Processing and Control*, vol. 79, pp. 104193, 2023. https://doi.org/10.1016/j.bspc.2022.104193

[30] Wang L., Wang H., Song Y., and Wang Q., "MCPL-based FT-LSTM: Medical Representation Learning-based Clinical Prediction Model for Time Series Events," *IEEE Access*, vol. 7, pp. 70253-70264, 2019. DOI:10.1109/ACCESS.2019.2919683

[31] Wang L., Zhang Z., and Hu T., "Effectiveness of LODS, OASIS, and SAPS II to Predict in-Hospital Mortality for Intensive Care Patients with ST Elevation Myocardial Infarction," *Scientific Reports*, vol. 11, no. 1, pp. 1-10, 2021. https://doi.org/10.1038/s41598-021-03397-3

[32] Wang S., McDermott M., Chauhan G., Ghassemi M., Hughes M., and Naumann T., "MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III," *in Proceedings of ACM Conference on Health, Inference, and Learning*, Toronto, pp. 222-235, 2019. http://dx.doi.org/10.1145/3368555.3384469

[33] Wu Y., Jiang M., Xu J., Zhi D., and Xu H., "Clinical Named Entity Recognition Using Deep Learning Models," *AMIA Annual Symposium Proceedings*, vol. 2017, pp. 1812-1819, 2018. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977567/

[34] Xie F., Zhou J., Lee J., Tan M., Li S., Rajnthern L., Chee M., Chakraborty B., Wong A., and Dagan A., "Benchmarking Emergency Department Prediction Models with Machine Learning and

Public Electronic Health Records," *Scientific Data*, vol. 9, no. 658, pp. 1-12, 2022. https://doi.org/10.1038/s41597-022-01782-9

[35] Yin Y. and Chou C., "A Novel Switching State-Space Model for Post-ICU Mortality Prediction and Survival Analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3587-3595, 2021. DOI:10.1109/JBHI.2021.3068357

[36] Zangmo K. and Khwannimit B., "Validating the APACHE IV Score in Predicting Length of Stay in the Intensive Care Unit Among Patients with Sepsis," *Scientific Reports*. vol. 13, no. 1, pp. 1-9, 2023. https://doi.org/10.1038/s41598-023-33173-4

Daphne Samuel obtained her master of engineering in CSE at Karunya University, India during 2012. Currently doing Ph.D. at Department of Computer Science and Engineering, CEG Campus, Anna University, Chennai.

Mary Anita Rajam obtained her Ph.D. during 2010 at College of Engineering, Anna University. Master of Engineering during 1997 in CSE at Government College of Engineering, Tirunelveli. Currently is the Professor, DCSE and Director, Centre for Cyber Security, CEG, Anna University, Chennai.