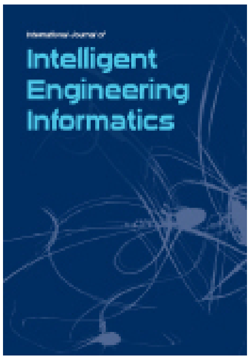


**International Journal of Intelligent Engineering Informatics**

ISSN online: 1758-8723 - ISSN print: 1758-8715
https://www.inderscience.com/ijiei

# Context parameters retrieval framework from electronic healthcare record through biomedical NLP for clinical support

Gaurav Paliwal, Aaquil Bunglowala, Pravesh Kanthed

**DOI:** 10.1504/IJIEI.2023.10054808

**Article History:**

| | |
|---|---|
| Received: | 17 November 2022 |
| Last revised: | 16 January 2023 |
| Accepted: | 18 January 2023 |
| Published online: | 03 May 2023 |

*Int. J. Intelligent Engineering Informatics, Vol. 11, No. 1, 2023*

# Context parameters retrieval framework from electronic healthcare record through biomedical NLP for clinical support

## Gaurav Paliwal* and Aaquil Bunglowala

School of Technology Management and Engineering,
SVKM's NMIMS,
Super Corridor Rd, Gandhi Nagar,
Indore, Madhya Pradesh 452005, India
Email: Gaurav.paliwal@nmims.edu
Email: Aaquil.Bunglowala@nmims.edu
*Corresponding author

## Pravesh Kanthed

Choithram Hospital and Research Centre,
14, Manik Bagh Rd, Choitram Compound,
Indore, Madhya Pradesh 452014, India
Email: praveshkanthed@gmail.com

**Abstract:** This paper presents architecture for extracting context features from electronic health records (EHR) to design a clinical support system using transfer learning for natural language processing (NLP). The system is trained to provide a supporting summary to the medical practioners on the basis of the ICD 9 codes and respective symptoms of the patient. The BioALBERT model has been trained over biomedical corpora and the proposed model uses improvised parameter sharing techniques and requires less physical memory. The theoretical analysis of the proposed system is supported by the experimental analysis. MIMIC-III database has been used to fine train the proposed models and to assess the efficiency and efficacy of the proposed work. This study introduced a context-aware approach for extracting useful context from EHR, which can be used to acquire a basic understanding of the treatment path.

**Keywords:** natural language processing; NLP; transfer learning; attention mechanisms; biomedical NLP; ALBERT; BioALBERT; electronic health records; EHR.

**Reference** to this paper should be made as follows: Paliwal, G., Bunglowala, A. and Kanthed, P. (2023) 'Context parameters retrieval framework from electronic healthcare record through biomedical NLP for clinical support', *Int. J. Intelligent Engineering Informatics*, Vol. 11, No. 1, pp.1–18.

**Biographical notes:** Gaurav Paliwal has 11 years of experience in teaching and research. He is currently pursuing his PhD in Healthcare. He holds a Master of Technology degree in Computer Engineering and Master of Business Administration in Information System Management. He has extensive research in mobile patient monitoring systems and wearable monitoring architectural

design, published in two books, two book chapters and various international journal and conference papers. His areas of expertise include knowledge extraction, clinical healthcare systems, natural language processing, deep learning and wearable sensor networks.

Aaquil Bunglowala has over 22 years of experience in industry and academics on various profiles, he has substantial experience in sophisticated research techniques and technologies in advanced logic and VLSI design. He has been actively involved in research. He has published more than 52 research papers in peer national/international journals and conferences. He has also contributed in research field by becoming reviewer in various international conferences and journals. He has organised many international and national conferences He has chaired many scientific sessions of national and international conferences.

Pravesh Kanthed is a Consultant and Head of Pain Management, associated with Choithram Hospital and Research Center, Indore. He deals with non-surgical to minimally invasive treatments of Spine Pain. He is the First Interventional Pain Physician of Central India and practicing for the last 11 years. Conditions treated at pain clinic includes back and leg pain, neck and arm pain, slip disc, sciatica, herniated disc, failed back surgery syndrome (spine pain after spine surgery), trigeminal/post herpetic neuralgia, headaches, joint pain, osteo-arthritis, fibromyalgia, diabetic neuropathy, cancer pain, post surgery scar pain, post operative pain management and providing pain relief during labour.

## 1 Introduction

EHR is the most powerful information pool of clinical data comprising of the medical and pathological reports, treatment, symptoms, and the respective clinical notes. Such information can aid in the development of a continuous information flow between patients and doctors. More importantly, comprehensive evaluations have shown that the clinical care quality can greatly be enhanced through the predictive analysis of the EHR (Meystre et al., 2008; Bui et al., 2022; Chen et al., 2021). Some freely available EHR resources, such as the Medical Information Mart for Intensive Care (MIMIC) database, and PubMed dataset provide a huge treasure of biomedical data to extract the insightful information for the clinical support. These datasets contain the healthcare data from patients admitted to a large tertiary care hospital's critical care unit. Structured data contains diagnostics and laboratory results, while unstructured data contains doctor progress notes and discharge summaries. The analysis and information retrieval are comparatively more challenging in case of unstructured data especially in case of biomedical field (Cohen and Hersh, 2005; Chifu et al., 2019; Chai et al., 2021).

One of the most important tasks in secondary evaluation of EHR is appropriately identifying the patient cohort under examination. Clinical narratives often provide very insightful and meaningful information for an appropriate diagnosis of medical situation of any patient. As a result, NLP techniques must be used to obtain and assess this information. The thorough analysis of this text through NLP and advance ML tools can be utilised to identify the clinical classification of the symptoms and the respective disease. It not only provides a better insight of health and disease, but also derives an accurate phenotyping which could be used to determine the risk of a person suffering from any disease (Stanfill et al., 2010; Tiwari et al., 2021). Furthermore, an accurate phenotyping of the patients through efficient software can reduce the time consumed by the doctors in evaluating the medical charts. Many researchers have used the NLP techniques to derive the intelligent predictive models for the phenotyping (Nigam, 2016). Considering the fact that completely rule-based predictive models require a lot of domain expertise, and it is difficult to ensure its accuracy in random situation, deep learning techniques for phenotyping advancements could be proved as a great tool for phenotyping. The manner in which information is derived from the EHR data makes interpretation and transmission of the data highly challenging. Compound clinical concepts are created by connecting the separate records together to derive a context which could be used to present a support to the medical practitioners and clinicians in case of emergency. These distinct records are combined in order to derive the appropriate meaning best suited and connected with the earlier entries in the healthcare records under several sections. When interpreting these hierarchically layered medical records, contexts are crucial. They undoubtedly aid in determining the severity, lack of discoveries, and clear depictions of medical disorders (Erdil, 2019).

Deep learning has showed promising results in various areas of healthcare, such as mortality prediction and patient identification. However, the issues in biomedical NLP like de-identification of notes, diagnosis of skin cancer, and detection of diabetic retinopathy, etc can still pose a tough challenge in applying deep learning for the disease prediction. The interpretability of these forecasts is very critical for healthcare applications since the results might have a direct impact on patient health decisions. Furthermore, clinicians with extensive prior knowledge and experience still want some applications to help them make some preliminary judgements in general scenario. As a result, models with greater interpretability are essential so that the doctors could have more confidence and control over their findings. While tremendous work has been done into understanding and developing understandable deep learning NLP models, their intricate interactions between inputs make them fundamentally less interpretable as compared to the conventional dictionary-based linear models (Huang et al., 2019; Li et al., 2018; Shickel et al., 2017).

Transfer learning has gained a lot of attraction in the recent years in the field of NLP due to its capability to utilise the knowledge obtained from one task to complete another related task. It lowers the need for fine-tuning datasets and enhances performance. Domain adaptation is a property of the transfer learning technique that is obtained through word vector mapping between similar words and similar vectors. The pre-training of the entire model over a data-rich task achieves general-purpose knowledge and abilities in transfer learning. Following that, this knowledge is transferred to the downstream tasks. The vast amount of text data available on the internet might be used to train the network using transfer learning. Generative pre-trained transformer (GPT), embeddings from language models (ELMo), bidirectional encoder representations from transformers (BERT) and its variants like ALBERT, RoBERTa have been proposed for NLP applications (Peters et al., 2018; Devlin et al., 2019; Lan et al., 2019; Jin et al., 2019, Mullenbach et al., 2018). All these models were shown to be quite effective for one set of applications but not for the other more diverse applications. Even the techniques, methodology, and practises used in these models differ, and as a result, their performance differs depending on the task. Therefore, for a more thorough knowledge of transfer learning, a unified and systematic approach is required. Also, the performance of the conventional pretrained learning models degrades for biomedical applications as they were trained over the general corpora like wordnet, Wikipedia, etc. The ineffective performance of these models over biomedical data has been resolved by biomedical NLP (BioNLP) researchers by training them on biological and clinical corpora (Yuan et al., 2021).

An architecture for extracting the context features to design a clinical support system using transfer learning for the natural language processing (NLP) over the biomedical data is presented in this paper. These factors can be utilised to interpret the patients' health state and extract necessary knowledge from the system in case of emergency. The large scale pretrained model, BioALBERT has been finely tuned for domain-specific adaptation over corpora from biomedical field. The algorithm presented in this paper has been fine-tuned to extract contextual information from an EHR system for clinical usage. The major contribution of the proposed work can be summarised as below:

- The derivation of a clinical support system using the contextual information in a clinical domain from the MIMIC-III dataset which could be of great use for the medical practitioners.
- Training of the intelligent support system to provide a supporting summary to the medical practioners on the basis of the ICD 9 codes and respective symptoms of the patient. The summary comprises of the accurate phenotyping, line of treatment, possible side effects, special care to take, etc for the patient with respective medical situation using the patient cohort.
- Fine tuning of the pretrained model with the contextual biomedical data to achieve the desired performance.
- The improvised parameter sharing techniques adds more value to the research work as it results into lesser physical memory requirement.
- The word piece embeddings through the sentence piece tokenisation for the fine-tuning for contextual summary generation also adds novelty to the proposed work.

We hypothesise that training biomedical variant of ALBERT, i.e., BioALBERT with biomedical corpora and further tuning it finely for the task of context aware summary generation can present an efficient support system for the clinical analysis in a wide variety of situations. The objective of the proposed work is to derive a framework for extracting the context features to design a clinical support system using transfer learning for the NLP over the biomedical data. These factors can be utilised to interpret the patients' health state and extract necessary knowledge from the system in case of emergency.

The rest of the paper is structured as follows: the second section of the paper is devoted to a review of relevant research that has recently been presented by a number of different researchers. Section 3 provides an architectural description of ALBERT and its biomedical variant. For the available dataset, Section 4 presents a BioALBERT-based context information retrieval system. The experimental investigation in Section 5 evaluates and discusses the performance of the proposed technique, and Section 6 concludes the work.

## 2 Related work

The ever-increasing growth in the healthcare sector due to the digitisation of data and the potential of artificial intelligence to extract the meaningful information from it have attracted many researchers to address variety of untouched aspects of medical paradigm. Recent development in the area of NLP has taken the capability of processing to the next level. Especially, Biomedical NLP has changed the complete spectrum of data analytics of medical sector, drug development process, sequence estimation of biological variations, etc. The application of advance language models (LMs) over the biomedical literature could not generate the encouraging results because of their training over general corpora. The limitation has been overcome by the BioNLP researchers by training these LMs on biological and clinical corpora and verified the respective performance over numerous BioNLP applications.

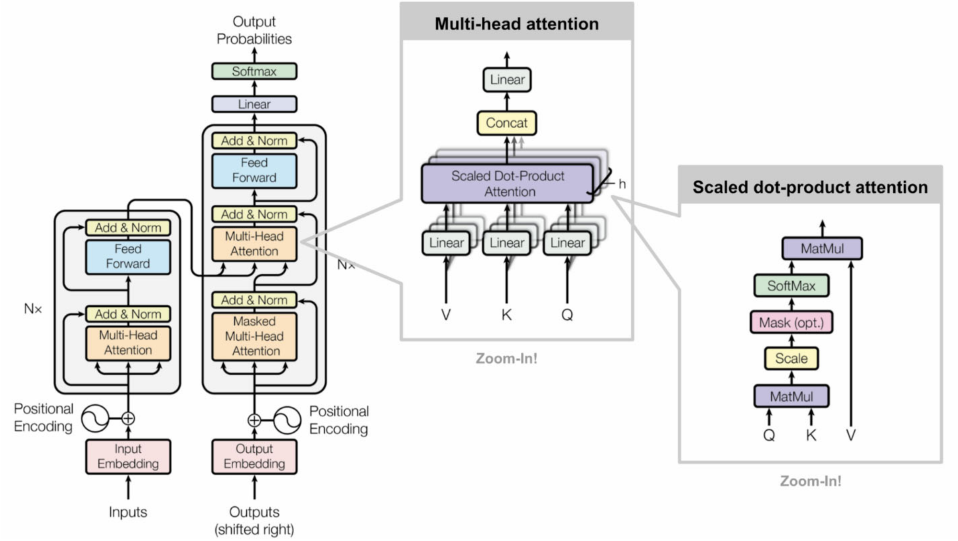
Jin et al. (2019) presented the variant of ELMo by training it using PubMed abstracts and named it as biomedical ELMo (BioELMo). The authors have retrieved the features of the biomedical data in terms of relational information among various entities. The application of BERT for training over scientific materials was proposed by Beltagy et al. (2019) and named the proposed model as Scientific BERT (SciBERT). On the same line, Si et al. (2019) have utilised the clinical notes as corpora for the training of BERT-based transfer learning models. They have targeted the problem of named-entity-recognition (NER) for biomedical data and attained and improved standard non-contextual and contextual word embedding. Peng et al. (2019) developed a benchmark named as biomedical language understanding evaluation (BLUE) score by using ten datasets to create five tasks which has been used by many researchers later to assess the correctness of their model. They proved that the performance of the BioNLP is enhanced when the LMs are trained over biomedical corpora. BioBERT presented by Lee et al. (2020) is the most commonly used LM for biomedical applications which utilises the PubMed and PubMed Central (PMC) corpus for pretraining. They have further explored the potential of the proposed model over three different tasks of BioNLP namely the relational extraction, NER and question answering (QA) through fine-tuning. PubMed articles have been used by Gu et al. (2020) who presented a novel LM and named it as PubMedBERT. They also verified the effectiveness and accuracy of the model with the LM trained over general corpora and also derived a domain-specific vocabulary. KeBioLM proposed by Yuan et al. (2021) which was a biological pre-trained LM, drew on knowledge from the unified medical language system (UMLS) database. KeBioLM was used to complete two BioNLP assignments and demonstrated the feasibility of developing a customised domain-specific LM.

Because all these pre-trained LMs use the BERT architecture, training takes a long time and demands a lot of computing power. Furthermore, all these LMs were demonstrated using only a few BioNLP tasks, hence their generalisability remains unknown. In NLP tasks, the performance of ALBERT is established to superior as compared to that of BERT (Naseem et al., 2021). All these researchers have touched upon various aspects of artificial intelligence and exploited the best of the characteristics of these LMs. However, the application of these LMs over the biomedical data to explore the depth of treasure of information is still not reached to the desired state. The major limitation of these models is the unavailability of the pre-trained models for the large biomedical data. It can be summarised that the application of the state-of-the-art transfer learning model to the biomedical field can revolutionise the complete paradigm of healthcare sector. The potential of advanced LM to extract the insight from the EHR can be of great use for generations to come.

## 3 Architectural framework of ALBERT

ALBERT is a lite variant of BERT which preserves the functionality of BERT but features much fewer parameters than typical BERT architecture. The parameter reduction is achieved through the augmentation of the factorised embedding parameterisation (FEP) technique with the subsequent technique of cross-layer parameter (CLP) sharing. The larger embedding matrix of vocabulary is reduced to two smaller matrices in FEP to decrease the dependence of the dimensions of hidden layer over the vocabulary embedding. Additionally, it provides the option to increase the hidden layer number while maintaining the same amount of parameters. On the other hand, the parameter efficiency is enhanced through CLP sharing technique by limiting the increase in the number of parameters even with the change in the depth of the network. It also presents a more regularised and stable training performance for the backbone BERT model. It has been found that an ALBERT model has utilised the parameters 18 times lesser than the conventional BERT model and still achieved a training performance which is 1.7 times faster than BERT.

**Figure 1** Classic transformer model (see online version for colours)



*Source:* Naseem et al. (2021)

The backbone framework of ALBERT is encoder-decoder-based model with multiple layer transformer architecture, as illustrated in fig 1. The encoder in this model is based on self-attention, but the decoder has attention on encoder outputs. It is made up of many blocks that are placed on top of each other. Each block comprises of a multi-head attention block and a Feedforward Network. To accommodate the information about the context of the data and word level embedding, hidden layer representations should be large which increases the number of parameters also. Typically, it is the product of number of tokens in the vocabulary ($V$) and the hidden layer size, i.e., $V * H$.

Considering $E$ as the size of the embedding after it has been factored using FEP, the parameters are drastically reduced to the order of $V * E + E * H$. Although the learning potential of the model can be improved by stacking the independent layers, it also dramatically increases their redundancy. This redundancy is addressed by ALBERT by sharing parameters between groups of layers. This presents a trade-off between the number of parameters and the number of layers. So, ALBERT is a small yet very helpful variation of BERT. It can increase the effectiveness of downstream language understanding activities while keeping the computational overhead to a minimum for a variety of applications.

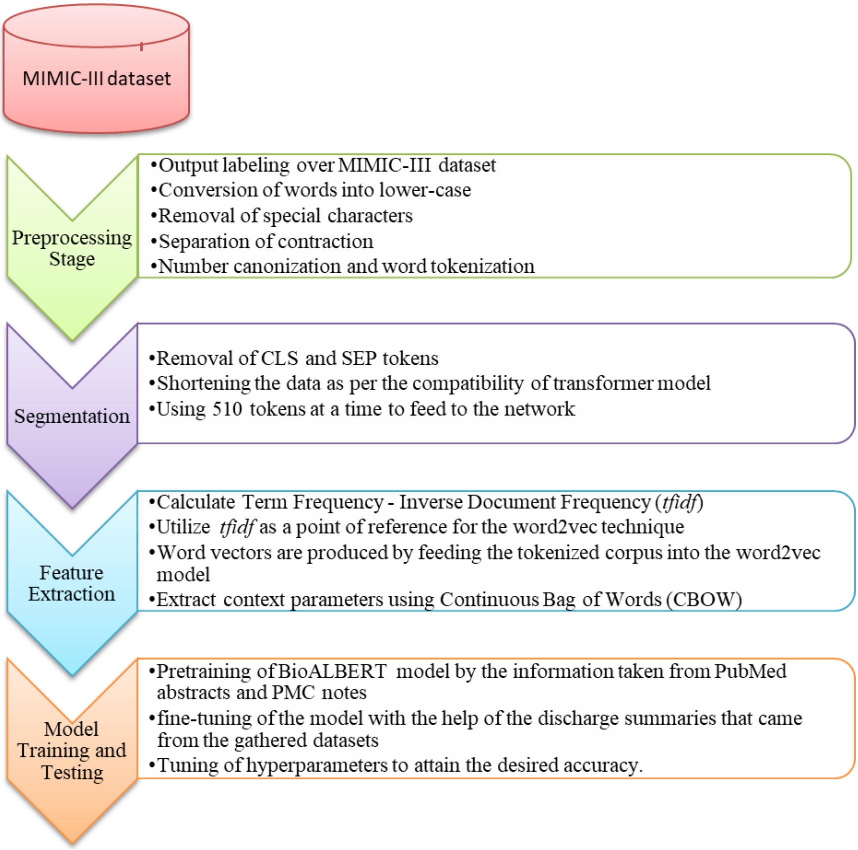
## 4 Proposed BioALBERT-based associated context parameters evaluation

This paper presents an architecture for extracting the context features to design a clinical support system using transfer learning for the NLP over the biomedical data. BioALBERT, has been used here to implement the transfer learning by fine tuning the large-scale pretrained LM over biomedical corpora. The whole work is majorly classified in four steps as shown in Figure 2. The steps are discussed in detail as follows.

### *4.1 Data preprocessing*

The dataset used in this work is MIMIC-III dataset which is a huge database of patients admitted to Beth Israel Deaconess Medical Center’s intensive care units in the time period of 2001 to 2012 (Johnson et al., 2016). The purpose of this research is to leverage unstructured data to uncover meaningful semantic information. Therefore, different files have been explored and processed in the proposed work to derive a compiled data of meaningful clinical information. The most important feature of the dataset is the discharge summaries which provided the cumulative textual information about the complete treatment plan of the patient. It is processed further through NLP techniques to extract the actual ground truth about the patients’ medical situation. The notes were sterilised in this stage by deleting any reference of class designations, misspelled words, special characters, repeated words, unidentified words, etc. The MIMIC dataset is preprocessed by performing the output labelling over MIMIC-III dataset, conversion of words into lower-case, removal of special characters, separation of contraction, number canonisation and word tokenisation. The complete dataset is further refined on the basis of the most frequently encountered ICD9 codes. And it has been refurbished under the categories of top 10 codes dataset and top 50 codes dataset. It has been observed that these two categories of datasets covered majority of the data. The tokenisation of the text of datasets under the discharge summaries features have been performed using the classification [CLS] token and separator [SEP] tokens. The start of the sentence is represented by [CLS] token and end of the sentences is represented by [SEP] token.

**Figure 2** Proposed methodology (see online version for colours)



## *4.2 Segmentation*

The technique of distinguishing various sentences among a bunch of words is known as sentence segmentation or sentence tokenisation. Sentence segmentation is carried out using the Spacy library, an NLP-designed library, with significantly greater accuracy. The segmentation of the sentences in the discharge summaries of the preprocessed dataset is performed using the CLS and SEP tokens and the existing tokenisation libraries. For transformer models, the real limit for the input layer is 510 tokens, down from 512 after removing the CLS and SEP tokens (Romanov and Shivade, 2018). Since the average length of the MIMIC-III discharge summaries is 1,947 tokens and only 11.67% of all documents are lengthier than 510 tokens, the data needs to be reduced so that it may be inserted into the transformer model. Despite the narrative style of the summaries, most of the documents follow a standard format: it usually starts with a chief complaint and proceed to a historical background portion where details regarding the patient's previous and present illnesses, as well as their socioeconomic and racial origins, are offered. The structure of diagnostic tests and relevant results is less consistent, and there are a few subsections that are more case-specific. The documentation includes a 'transitory hospital course' portion in the centre that describes the patient's time in the intensive care unit, and a 'discharge instructions' section at the end that specifies the requirements for the patient's release from the hospital. The material up to the point of the transitory hospital course will be eliminated, and the rest of the text will be used in order until all 510 tokens. If the transitory hospital course was absent from 822 papers, those summaries were not included.

## 4.3 Feature extraction

At this stage in the work, significant features are extracted from the segmented data namely term frequency-inverse document frequency (*tfidf*) (Sun et al., 2019), which is widely utilised in the field of NLP. After that, the *tfidf* is utilised as a point of reference for the word2vec technique, which is used to extract the more in-depth properties. The frequency of a word's appearance in a document (which is referred to as *tf*) and the rare or frequent appearance of the word in a document (which is referred to as *idf*) are the two statistics that are used in the *tfidf* in order to determine the significance of a word in a document. The total number of documents, represented by *nd*, and the number of documents containing the word *w*, represented by *df*(*d*, *w*), are used to determine the parameter *idf*, as follows (Huddar et al., 2020):

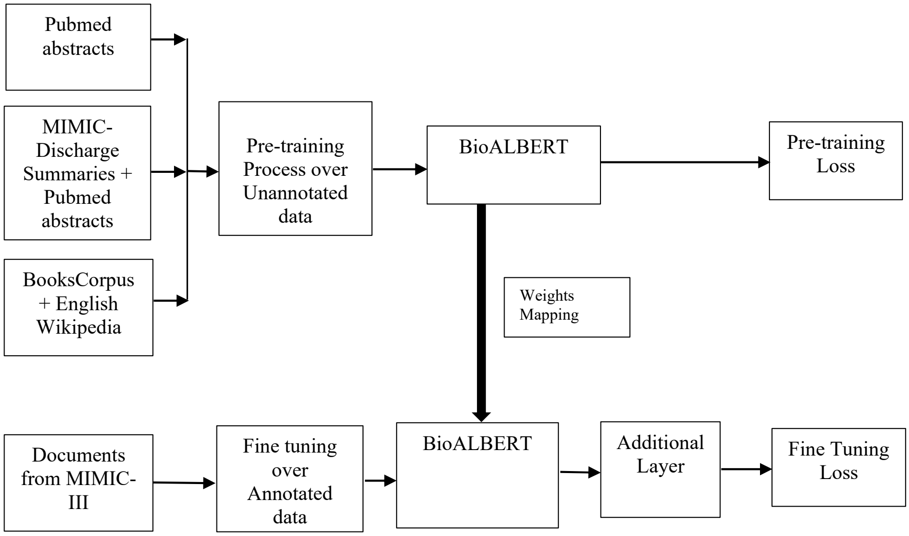
$$idf(w) = \log \frac{nd}{df(d, w)} + 1$$

To compute the *tfidf*, the statistical distribution of *tf* throughout the dataset is used, which is made possible by the tokenisation of the training text. This allows for the computation of the *tfidf*. After that, the relevant *idf* was multiplied by each word in the list. A tokenised corpus serves as the input for the *word2vec* model, which enables the generation of word vectors using the model. In order to discover the words that are adjacent to the given one in order to extract the context, we made use of an architecture called continuous bag of words (CBOW). Its structure is very similar to that of a neural network, and it takes projections as its inputs. In an effort to make the time series easier to understand, this method does not include the traditional nonlinear hidden layer that is typically seen in the output. In addition to this, the context of the word acts as an input, and the data that is contained in the projection layer is the same for each and every word. The MIMIC-III text notes corpora were utilised so that our word2vec model could be trained (Khatir and Nait-Bahloul, 2019).

## 4.4 Model training and testing

Standard LMs have demonstrated their potential for a wide range of NLP issues, however there are substantial pitfalls associated with utilising them for EMR. A lack of training data, the use of acronyms, the possibility that a single entity could be a member of more than one entity type, contextual dependency, and other factors are among these challenges. As a consequence of this, and as was mentioned in the introduction, the most advanced EMR models make use of context-dependent language models that are trained on biomedical corpora and are implemented using transformers. In this particular investigation, the more advanced and speedier LM known as the bioALBERT transformer model was utilised. Because of the model's capacity to adapt to the context in which it is embedded, it is possible to circumvent some of the challenges that are inherent in traditional domain-specific language models (Mikolov et al., 2013). The complete end to end training procedure is shown in Figure 3.

**Figure 3** End-to-end training procedure



To improve the performance of BioALBERT in terms of its ability to learn, CLP sharing was utilised throughout the training process, which took place on a large dataset that was derived from biomedical corpora. The knowledge that was learnt in the first block may be used by the future blocks to learn the model's parameters owing to CLP sharing. This reduces the requirement for every layer of the model's parameters to be individually learned. In addition, contextual information is complemented by a technique called sentence-order-prediction (SOP), which evaluates the coherence loss of each sentence and works to reduce it as much as possible during training. This method provides a more in-depth understanding of the model as well as a more accurate depiction of it by making use of the training set to produce indiscriminate pairs of phrases. The model has been FEP-enhanced to split the massive vocabulary embedding matrix into two smaller matrices in order to lessen the dependence of the hidden layer dimensions on the vocabulary embedding.

Because, according to the transformer model, the embedding size is directly proportional to the depth of the hidden layers, the reasoning for this is as follows: it makes it possible for us to maintain the same amount of settings for word embedding while simultaneously expanding the number of hidden layers. The CLP sharing technique, on the other hand, places constraints on the expansion of the number of parameters in relation to the depth of the network. It enhances the consistency and dependability of the training results, as well as the efficiency of the parameters, all without compromising the efficacy of BERT. The vocabulary embedding parameter can have a bigger value assigned to it without necessitating an equivalent increase in the number of hidden layers being used. The next passage contains further details concerning the initial training of BioALBERT as well as the subsequent tuning that took place.

### *4.4.1 BioALBERT's pre-training*

The model has already been trained with the help of information taken from PMC notes and PubMed abstracts, which together include a total of 13.5 billion and 4.5 billion words, respectively. The text had to be further processed into a more structured manner in order to use the material from PMC and PubMed for the pre-training of the model. This was necessary in order to use the content. This raw text goes through a number of standard processing steps, including the elimination of blank lines and the formation of paragraphs, the deletion of sentences that are less than 20 characters long, and the addition of blank lines in the middle of each individual document. The data was first handled as a sentence text since tokenisation relies on sentence embedding's. This was done so that tokenisation could work properly. The fact that each sentence in the input text document is represented by a line in the document is the reason why there is a blank line between each document. There was a 512-character limit on the length of each declaration. The BioALBERT model was pretrained with this revised dataset so that it would not need to use the predefined vocabulary of the ALBERT model.

### *4.4.2 BioALBERT fine-tuning*

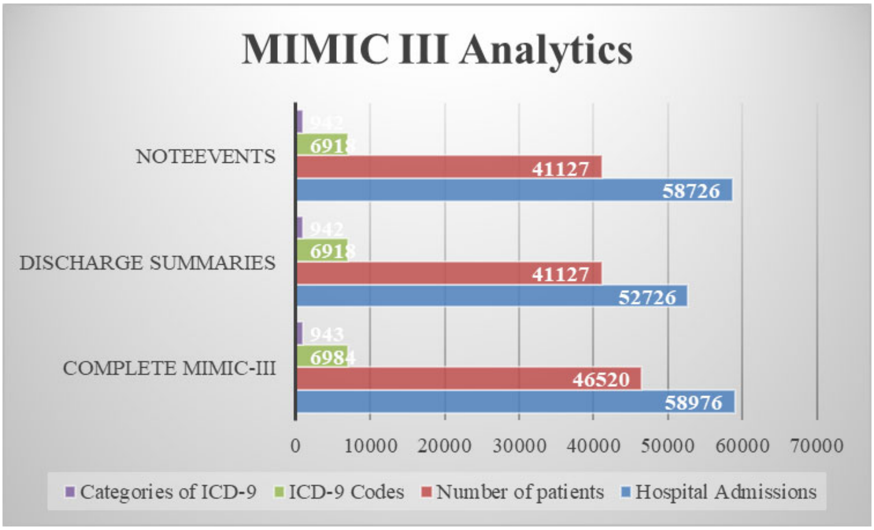
In this part, we will describe ways in which the contextual summary task that BioALBERT performs can be improved. The method involves selecting particular words from within a sentence so that those words can be dealt with as independent entities: racial composition, pathological results and HADM_ID. The proposed model was fine-tuned with the help of the discharge summaries that came from the gathered datasets. The objective is to become familiar with and make educated guesses regarding the nouns and keywords that are most significant in the field of biomedicine. As part of this assignment, you will compile a summary of the context that may be applied in clinical settings by using the symptom information and ICD 9 codes that have been provided. In contrast to pre-training, fine-tuning is straightforward and does not demand nearly as much cognitive capacity from the user.

The proposed approach makes use of improvised mechanisms for sharing parameters and utilises actual physical memory in a very limited capacity. The process of tokenising sentences into smaller chunks makes it easier to create word embeddings, which can then be utilised to generate more accurate contextual summaries. The one-of-a-kind dataset is put to use in the process of developing fine-tuning tasks for each of the models that have already been trained. The weights of the model were determined by using the pre-trained model that was constructed in advance. Before carrying out the fine-tuning process, our training was carried out using lowercase texts, with a learning rate of $1 \times 10^{-5}$, batch size of 32, and a total of 5,336 steps. In order to train the first version of the BioALBERT model, the Tensorflow TPUv3-8 processing unit was utilised. Most of the hyper-parameters have been left in the same states as they were in the traditional ALBERT model. The Adam optimiser was used at certain predefined times in time to evaluate the model using the evaluation checkpoint. The final prediction was arrived at by calculating it with the use of the holdout development dataset and the model that performed the best. During the course of the experiments, PyTorch-transformers that made use of BERT and XLNet implementations were utilised. The entirety of the training process, beginning with its inception and continuing all the way until its completion is depicted in Figure 2. Modifications were made to the models without the layers being frozen. Adam (Kingma and Ba, 2014) was used as the optimiser in the early stages of the study, but subsequently, researchers switched to the layer wise adaptive large batch (LAMB) optimiser (Yang et al., 2019; Mittal and Nagpal, 2021), which resulted in a noticeably shorter training period. When determining which hyperparameters to use and how to optimise them, we took into account the overall performance of the development set. In most cases, the results of attempting to apply the Transformer model with a variety of learning rates were unsatisfactory, which advocates that the model is quite sensitive to the parameters it is given. The best results were achieved by training at a pace of 7 × $10^{-4}$ or 6 × $10^{-6}$. In order to rank the codes according to their level of relative confidence, the binary cross entropy logits were utilised.

## 5 Results and discussion

The MIMIC-III dataset was used in an experimental study (Johnson et al., 2016) that was conducted. We used ICD-9 codes from the dataset that had been meticulously annotated in order to serve as labels. The complete MIMIC-III dataset consists of 52,724 discharge summaries and contains 8,922 unique classifications that were applied to those summaries. It includes summaries of 8,066 discharges that were used for training, 1,573 discharges that were used for validation, and 1,729 discharges that were used for testing in relation to the experimental inquiry. Through this line of inquiry, we want to acquire the skills necessary to derive meaningful semantic information from raw data. In order to accomplish this, we relied solely on the free-text clinic notes included in the dataset, more specifically the noteevents table. In addition, the discharge summaries category stood out from the others since, in contrast to the others, it included genuine ground facts as well as free text. The total number of patients, hospital admissions, ICD-9 codes, and ICD-9 categories are presented in Figure 4. MIMIC-III contains the entire dataset, in contrast to noteevents and discharge summaries, which only include the subgroups that are of interest.

**Figure 4** MIMIC III dataset analytics (see online version for colours)



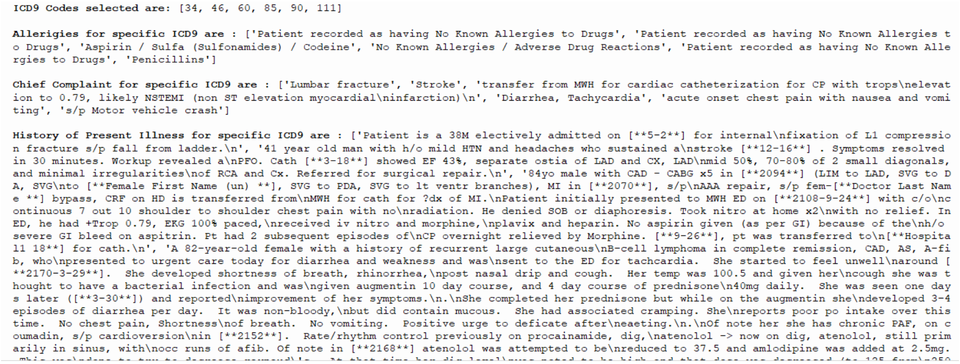
For the base model of BioALBERT, a 1,024 training batch size was used; however, due to constraints associated with the available computational resource, the size of the training batch was reduced to 256. Table 1 lists the hyperparameters used in the training process.

**Table 1** Pre-training parameters

| *Parameters* | *Description or values* | *Parameters* | *Description or values* |
|---|---|---|---|
| Baseline model | ALBERT | Number of layers | 12 |
| Optimisation technique | LAMB | Hidden layers | 768 |
| Kernel | GeLU | Size of vocabulary | 30,000 |
| Maximum length of sentence | 512 | Warm-up steps | 3,125 |
| Number of attention heads | 12 | Evaluation batch size | 16 |
| Training batch size | 1,024 | Size of embedding | 128 |

When compared to other baseline models, BioALBERT's computing requirements were not much higher, and fine-tuning only required lesser computation as compared to the process of pre-training. The performance of BioALBERT is further enhanced due to its lesser physical memory requirement and spontaneous parameter sharing approaches. The training process is faster because of sentence piece tokenisation to learn the word embeddings. The weights of the pre-trained BioALBERT are utilised during the fine tuning of the model. AdamW optimiser is used here with a learning rate of 0.00001 and a batch size of 32. For all jobs, we limited sentence length to 128 characters and capitalised all words. Finally, during the fine-tuning procedure, we employed 512 warm-up steps to fine-tune our pre-trained models for 10k training steps.

**Figure 5** User interface-1 with supporting context

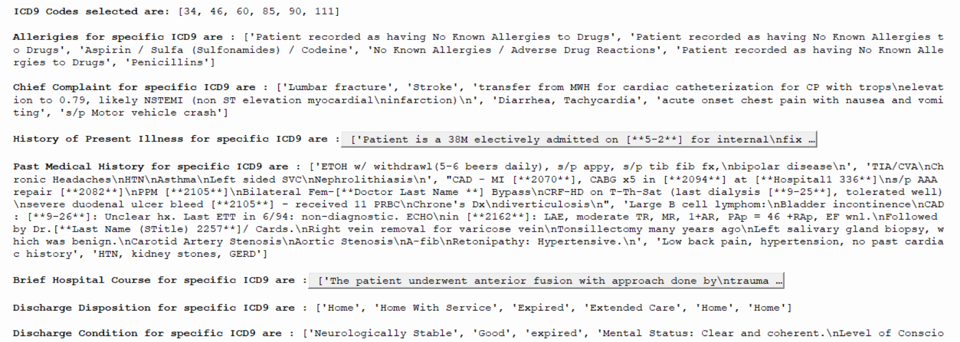


```
ICD9 Codes selected are: [34, 46, 60, 85, 90, 111]

Allerigies for specific ICD9 are : ['Patient recorded as having No Known Allergies to Drugs', 'Patient recorded as having No Known Allergies t
o Drugs', 'Aspirin / Sulfa (Sulfonamides) / Codeine', 'No Known Allergies / Adverse Drug Reactions', 'Patient recorded as having No Known Alle
rgies to Drugs', 'Penicillins']

Chief Complaint for specific ICD9 are : ['Lumbar fracture', 'Stroke', 'transfer from MWH for cardiac catheterization for CP with trops\nelevat
ion to 0.79, likely NSTEMI (non ST elevation myocardial\ninfarction)\n', 'Diarrhea, Tachycardia', 'acute onset chest pain with nausea and vomi
ting', 's/p Motor vehicle crash']

History of Present Illness for specific ICD9 are : ['Patient is a 38M electively admitted on [**5-2**] for internal\nfixation of L1 compressio
n fracture s/p fall from ladder.\n', '41 year old man with h/o mild HTN and headaches who sustained a\nstroke [**12-16**] . Symptoms resolved
in 30 minutes. Workup revealed a\nPFO. Cath [**3-18**] showed EF 43%, separate ostia of LAD and CX, LAD\nmid 50%, 70-80% of 2 small diagonals,
and minimal irregularities\nof RCA and Cx. Referred for surgical repair.\n', '84yo male with CAD - CABG x5 in [**2094**] (LIM to LAD, SVG to D
A, SVG\nto [**Female First Name (un) **], SVG to PDA, SVG to lt ventr branches), MI in [**2070**], s/p\nAAA repair, s/p fem-[**Doctor Last Nam
e **] bypass, CRF on HD is transferred from\nMWH for cath for ?dx of MI.\nPatient initially presented to MWH ED on [**2108-9-24**] with c/o\nc
ontinuous 7 out 10 shoulder to shoulder chest pain with no\nradiation. He denied SOB or diaphoresis. Took nitro at home x2\nwith no relief. In
ED, he had +Trop 0.79, EKG 100% paced,\nreceived iv nitro and morphine,\nplavix and heparin. No aspirin given (as per GI) because of the\nh/o
severe GI bleed on aspitrin. Pt had 2 subsequent episodes of\nCP overnight relieved by Morphine. [**9-26**], pt was transferred to\n[**Hospita
l1 18**] for cath.\n', 'A 82-year-old female with a history of recurrent large cutaneous\nB-cell lymphoma in complete remission, CAD, AS, A-fi
b, who\npresented to urgent care today for diarrhea and weakness and was\nsent to the ED for tachcardia.  She started to feel unwell\naround [
**2170-3-29**].  She developed shortness of breath, rhinorrhea,\npost nasal drip and cough.  Her temp was 100.5 and given her\ncough she was t
hought to have a bacterial infection and was\ngiven augmentin 10 day course, and 4 day course of prednisone\n40mg daily.  She was seen one day
s later ([**3-30**]) and reported\nimprovement of her symptoms.\n.\nShe completed her prednisone but while on the augmentin she\ndeveloped 3-4
episodes of diarrhea per day.  It was non-bloody,\nbut did contain mucous.  She had associated cramping. She\nreports poor po intake over this
time.  No chest pain, Shortness\nof breath.  No vomiting.  Positive urge to deficate after\neaeting.\n.\nOf note her she has chronic PAF, on c
oumadin, s/p cardioversion\nin [**2152**].  Rate/rhythm control previously on procainamide, dig,\natenolol -> now on dig, atenolol, still prim
arily in sinus, with\nocc runs of afib. Of note in [**2168**] atenolol was attempted to be\nreduced to 37.5 and amlodipine was added at 2.5mg.
```

A user interface has also been designed in the proposed work which takes the respective ICD-9 codes as input and generates the contextual short summary of the patients suffering from the similar symptoms. This summary can be used by the clinicians or medical practitioners in the emergency cases. The treatment provided to the patients with similar symptoms and respective results along with the mortality status can be of great value to the medical fraternity. Few screenshots of the designed user interface are shown in Figures 5 and 6. It is shown in the screenshots of the user interface that when the ICD-

9 codes corresponding to respective situation of the patient are entered, it generates the supporting context of the situation which is derived from the similar types of patients with same situation and their respective treatment plans from the dataset. The context support includes the possible allergies, symptoms, clinical histories, treatment plan and the mortality status of the earlier patients. This summary can be followed by the medical practitioners and clinicians to decide the roadmap for treatment in emergency situations. It can also be used for the further research to evaluate the impact of any medical treatment for the future as well.

**Figure 6** User interface-2 with supporting context



```
ICD9 Codes selected are: [34, 46, 60, 85, 90, 111]

Allerigies for specific ICD9 are : ['Patient recorded as having No Known Allergies to Drugs', 'Patient recorded as having No Known Allergies t
o Drugs', 'Aspirin / Sulfa (Sulfonamides) / Codeine', 'No Known Allergies / Adverse Drug Reactions', 'Patient recorded as having No Known Alle
rgies to Drugs', 'Penicillins']

Chief Complaint for specific ICD9 are : ['Lumbar fracture', 'Stroke', 'transfer from MWH for cardiac catheterization for CP with trops\nelevat
ion to 0.79, likely NSTEMI (non ST elevation myocardial\ninfarction)\n', 'Diarrhea, Tachycardia', 'acute onset chest pain with nausea and vomi
ting', 's/p Motor vehicle crash']

History of Present Illness for specific ICD9 are : ['Patient is a 38M electively admitted on [**5-2**] for internal\nfix …

Past Medical History for specific ICD9 are : ['ETOH w/ withdrawl(5-6 beers daily), s/p appy, s/p tib fib fx,\nbipolar disease\n', 'TIA/CVA\nCh
ronic Headaches\nHTN\nAsthma\nLeft sided SVC\nNephrolithiasis\n', "CAD - MI [**2070**], CABG x5 in [**2094**] at [**Hospital1 336**]\ns/p AAA
repair [**2082**]\nPPM [**2105**]\nBilateral Fem-[**Doctor Last Name **] Bypass\nCRF-HD on T-Th-Sat (last dialysis [**9-25**], tolerated well)
\nsevere duodenal ulcer bleed [**2105**] - received 11 PRBC\nChrone's Dx\ndiverticulosis\n", 'Large B cell lymphom:\nBladder incontinence\nCAD
: [**9-26**]: Unclear hx. Last ETT in 6/94: non-diagnostic. ECHO\nin [**2162**]: LAE, moderate TR, MR, 1+AR, PAp = 46 +RAp, EF wnl.\nFollowed
by Dr.[**Last Name (STitle) 2257**]/ Cards.\nRight vein removal for varicose vein\nTonsillectomy many years ago\nLeft salivary gland biopsy, w
hich was benign.\nCarotid Artery Stenosis\nAortic Stenosis\nA-fib\nRetonipathy: Hypertensive.\n', 'Low back pain, hypertension, no past cardia
c history', 'HTN, kidney stones, GERD']

Brief Hospital Course for specific ICD9 are : ['The patient underwent anterior fusion with approach done by\ntrauma …

Discharge Disposition for specific ICD9 are : ['Home', 'Home With Service', 'Expired', 'Extended Care', 'Home', 'Home']

Discharge Condition for specific ICD9 are : ['Neurologically Stable', 'Good', 'expired', 'Mental Status: Clear and coherent.\nLevel of Conscio
```

The training performance in terms of cost function is shown in Figure 7, which presents the optimal tuning of the proposed model. Figure 8 shows the variation of estimation error and the learning characteristics is reflected from converging nature of the error.

**Figure 7** Cost function v/s time (see online version for colours)

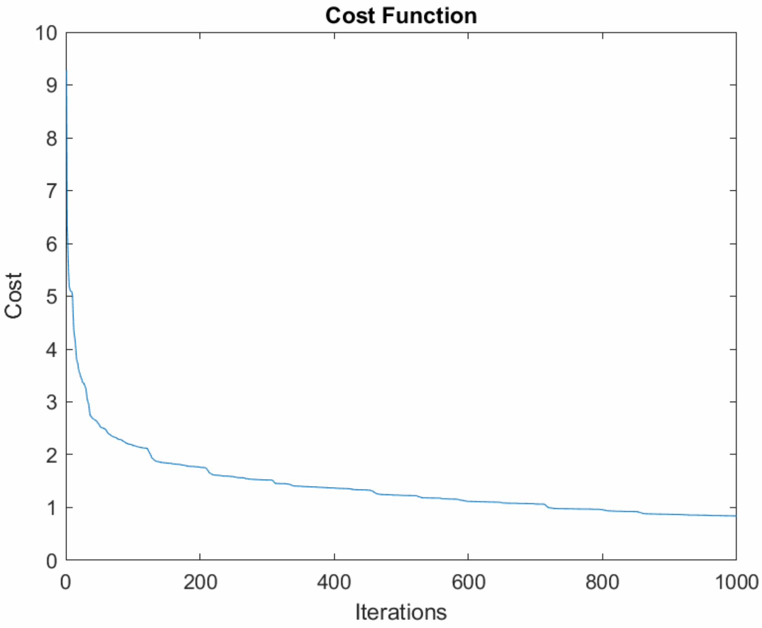


**Figure 8** Estimation error v/s time (see online version for colours)

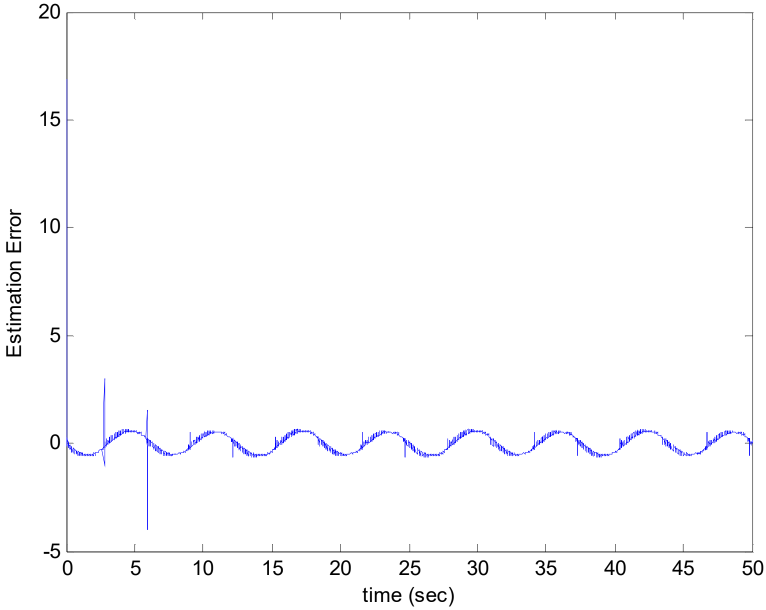


**Figure 9** Modelling accuracy v/s number of iterations (see online version for colours)

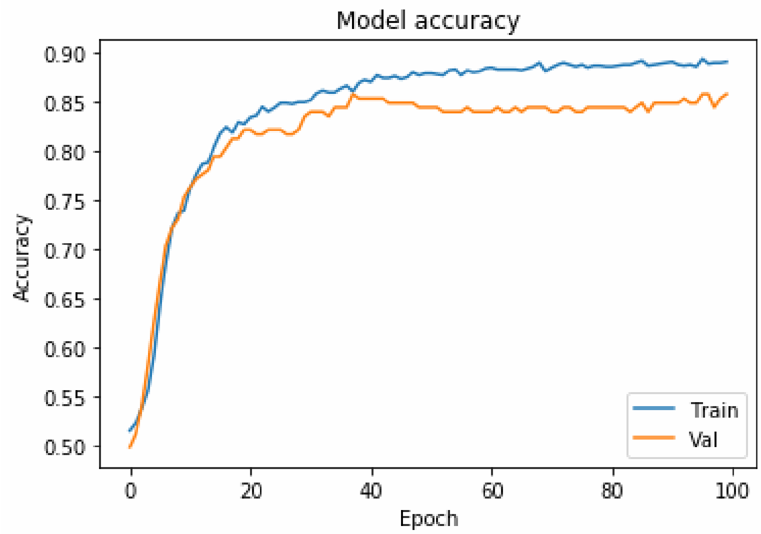
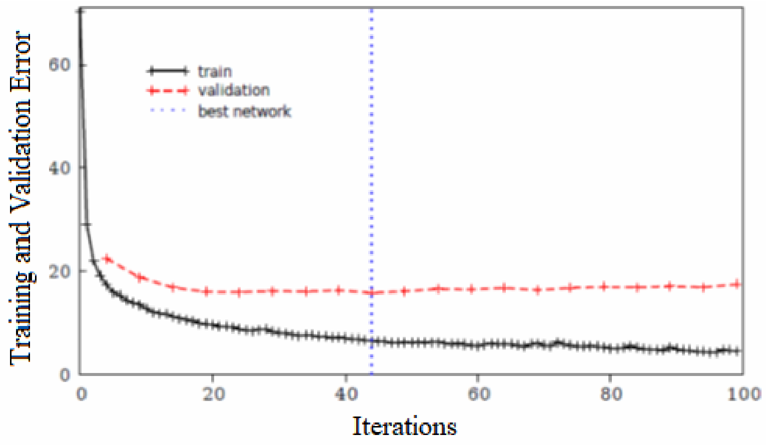


Figure 9 and 10 show the modelling accuracy and error during the training and validation over time for the training run on the online database.

**Figure 10** Training and validation error wrt number of iterations (see online version for colours)



### 5.1 *Discussion*

While some researchers have focused on using deep learning to automatically assign ICD-9 codes, others have stressed the importance of retrieving usable information from unstructured biomedical data. Because there are no common comparison measurements and attributes, it is impossible to compare existing research on such jobs. This research is noteworthy because it presented the context aware system to extract the insightful context from the biomedical data. This contextual information can be utilised by the medical practitioners in emergency situation to get the preliminary idea about the line of treatment. The tentative symptoms or tentative ICD-9 codes can be entered as input to this system and the proposed model will generate the suggestions on the basis of the treatment and outcome from past records from the dataset with similar medical situation. It is not necessary that the clinician or medical practitioners (with variable expertise and experience) should follow the exact treatment as suggested by the proposed system, but a suggestive insight can be of great help to them.

## 6 Conclusions

The use of transfer learning for biomedical NLP to generate the context-based EHR system is presented in this study. In this study, we used the BioALBERT model, which was trained on biomedical corpora. It has been fine-tuned to extract context from an EHR system for clinical applications. The suggested model relies on improvised parameter sharing approaches and has a smaller physical memory footprint. It uses sentence piece tokenisation to implement word piece embeddings for fine-tuning for contextual summary creation. The experimental examination of the suggested system backs up the theoretical analysis. The MIMIC-III database was used to fine-tune the proposed models and analyse the proposed work's efficiency and efficacy. This study is notable because it introduced a context-aware approach for extracting useful context from EHR. Medical professionals may utilise this contextual knowledge in a broad range of scenarios to gain a rudimentary idea of the treatment course.

## References

Beltagy, I., Lo, K. and Cohan, A. (2019) 'SCIBERT: a pretrained language model for scientific text', *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp.3606–3611.

Bui, QT., Ngo, M.P., Snasel, V. et al. (2022) 'The sequence of neutrosophic soft sets and a decision-making problem in medical diagnosis', *International Journal of Fuzzy System*, Vol. 24, No. 4, pp.2036–2053.

Chai, J.S., Selvachandran, G., Smarandache, F. et al. (2021) 'New similarity measures for single-valued neutrosophic sets with applications in pattern recognition and medical diagnosis problems', *Complex Intelligent System*, Vol. 7, No. 2, pp.703–723.

Chen, R.J., Lu, M.Y., Chen, T.Y. et al. (2021) 'Synthetic data in machine learning for medicine and healthcare', *National Biomedical Engineering*, Vol. 5, No. 6, pp.493–497.

Chifu, V.R., Chifu, E.S., Pop, C.B., Salomie, I. and Lupu, M. (2019) 'Evolutionary-based method for risk stratification of diabetic patients', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 1, pp.37–60.

Cohen, A.M. and Hersh, W.R. (2005) 'A survey of current work in biomedical text mining', *Briefings in Bioinformatics*, Vol. 6, No. 1, pp.57–71.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2019) 'Bert: pre-training of deep bidirectional transformers for language understanding', *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.4171–4186.

Erdil, D.C. (2019) 'A comparison of health informatics education in the USA', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 4, pp.366–383.

Gu, Y. et al. (2020) *Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing*, arXiv preprint arXiv:2007.15779.

Huang, J., Osorio, C. and Wicent, L. (2019) 'An empirical evaluation of deep learning for ICD-9 code assignment using MIMIC-III clinical notes', *Computer Methods and Programs in Biomedicine*, Vol. 177, pp.141–153.

Huddar, M.G., Sannakki, S.S. and Rajpurohit, V.S. (2020) 'Attention-based word-level contextual feature extraction and cross-modality fusion for sentiment analysis and emotion classification', *International Journal of Intelligent Engineering Informatics*, Vol. 8, No. 1, pp.1–18.

Jin, Q., Dhingra, B., Cohen, W. and Lu, X. (2019) 'Probing biomedical embeddings from language models', *Proceedings of the 3rd Workshop on Evaluating Vector Space Representations for NLP*, pp.82–89.

Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, H., Wei, L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A. and Mark, R.G. (2016) 'Mimic-III, a freely accessible critical care database', *Scientific Data*, Vol. 3, No. 1, p.160035.

Khatir, N. and Nait-Bahloul, S. (2019) 'New weighted clustering ensemble based on external index and subspace attributes partitions for large features datasets', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 4, pp.323–345.

Kingma, D.P. and Ba, J. (2014) *Adam: A Method for Stochastic Optimization*, arXiv preprint arXiv:1412.6980.

Lan, Z. et al. (2019) 'Albert: a lite BERT for self-supervised learning of language representations', *International Conference on Learning Representations*, pp.1–17.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H. and Kang, J. (2020) 'BioBERT: a pre-trained biomedical language representation model for biomedical text mining', *Bioinformatics*, Vol. 36, No. 4, pp.1234–1240.

Li, F., Liu, W. and Yu, H. (2018) 'Extraction of information related to adverse drug events from electronic health record notes: design of an end-to-end model based on deep learning', *JMIR Medical Informatics*, Vol. 6, No. 4, p.e12159.

Meystre, S.M., Savova, G.K., Schuler, K.C. and Hurdle, J.F. (2008) 'Extracting information from textual documents in the electronic health record: a review of recent research', *Yearbook of Medical Informatics*, Vol. 17, No. 1, pp.128–144.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J. (2013) 'Distributed representations of words and phrases and their compositionality', *Advances in Neural Information Processing Systems*, Vol. 26, pp.3111–3119.

Mittal, S. and Nagpal, C.K. (2021) 'Reinforcement learning based predictive analytics framework for survival in stock market', *International Journal of Intelligent Engineering Informatics*, Vol. 9, No. 3, pp.294–327.

Mullenbach, J., Wiegreffe, S., Duke, J., Sun, J. and Eisenstein, J. (2018) 'Explainable prediction of medical codes from clinical text', *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, New Orleans, Louisiana, Vol. 1 (Long Papers), pp.1101–1111.

Naseem, U., Khushi, M., Reddy, V., Rajendran, S., Razzak, I. and Kim, J. (2021) 'BioALBERT: a simple and effective pre-trained language model for biomedical named entity recognition', *International Joint Conference on Neural Networks (IJCNN)*, pp.1–7.

Nigam, P. (2016) *Applying Deep Learning to ICD-9 Multi-Label Classification from Medical Records*, Technical report, Stanford University.

Peng, Y., Yan, S. and Lu, Z. (2019) 'Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets', *Proceedings of the 18th BioNLP Workshop and Shared Task*, pp.58–65.

Peters, M. et al. (2018) 'Deep contextualized word representations', *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *Association for Computational Linguistics*, Vol. 1 (Long Papers), pp.2227–2237.

Romanov, A. and Shivade, C. (2018) 'Lessons from natural language inference in the clinical domain', *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp.1586–1596.

Shickel, B., Tighe, P.J., Bihorac, A. and Rashidi, P. (2017) 'Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis', *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, No. 5, pp.1589–1604.

Si, Y., Wang, J., Xu, H. and Roberts, K. (2019) 'Enhancing clinical concept extraction with contextual embeddings', *Journal of American Medical Informatics Association*, Vol. 26, No. 11, pp.1297–1304.

Stanfill, M.H., Williams, M., Fenton, S.H., Jenders, R.A. and Hersh, W.R. (2010) 'A systematic literature review of automated clinical coding and classification systems', *Journal of the American Medical Informatics Association: JAMIA*, Vol. 17, No. 6, pp.646–651.

Sun, C., Qiu, X., Xu, Y. and Xuang, X. (2019) 'How to fine-tune BERT for text classification?', *China National Conference on Chinese Computational Linguistics*, pp.194–206.

Tiwari, V., Verma, L.K., Sharma, P., Jain, R. and Nagrath, P. (2021) 'Neural network and NLP based chatbot for answering COVID-19 queries', *International Journal of Intelligent Engineering Informatics*, Vol. 9, No. 2, pp.161–175.

Yang, Y. et al. (2019) *Large Batch Optimization for Deep Learning: Training BERT in 76 Minutes*, arXiv:1904.00962 [cs.LG].

Yuan, Z., Liu, Y., Tan, C., Huang, S. and Huang, F. (2021) *Improving Biomedical Pretrained Language Models with Knowledge*, arXiv preprint arXiv:2104.10344, pp.1–11.