Glaucoma Surgery Outcome Prediction Using Progress Notes: A Comparative Study

Stanford CS224N Custom Project

Sarvesh R. Babu
Stanford University
sarveshb@stanford.edu

Samuel Barry
Stanford University
sambarry@stanford.edu

Abstract

We aimed to predict glaucoma surgery outcomes of patients who underwent a surgery at Stanford using pre-operative data. This is a critical problem in the field of ophthalmology as glaucoma is one of the leading cause of blindness worldwide. We developed various BERT-based multi-modal fusion models (sum, learnable sum, multiply, concatenate, SVD, multi-layer concatenation) leveraging both EHR structured data and progress notes. We used these architectures to develop a model that combines BERT's output into a recurrent neural networks (RNN) and implemented time embeddings (time2Vec) to exploit the temporal nature of the relationship between surgery outcome prediction and physician notes. We benchmarked the different methods and highlighted that pre-training greatly increases performance while the best-performing model is the late concatenation fusion model. Our contribution showed how we could extract some predictive power from these notes but also how current SOTA methods struggle with tasks humans also have a hard time with, such as this one.

1 Key Information to include

- External mentor: Sophia Ying Wang (Assistant Professor, Ophthalmology @ Stanford Medicine)
- External Collaborators: N/A
- Sharing project: N/A

2 Introduction

Motivation: Glaucoma is a leading cause of irreversible blindness worldwide. The estimated prevalence of glaucoma is rising, from an estimated 76 million in 2020 to 111.8 million in 2040 Tham et al. (2014). With surgery being the only viable long-term cure for patients, predicting if an operation is going to be a success or not is particularly important, yet unaddressed by the research community because of the difficulty of the task and the scarcity of data available for this particular application. Predictive models in ophthalmology have typically relied on testing data, such as retinal nerve fiber layer optical coherence tomography or visual fields (VFs). However, it has been a challenge to incorporate a patient's clinical history, which typically resides within the patient health record, into these predictive models. The adoption of electronic health records (EHRs) has presented an opportunity to develop machine learning and deep-learning models based on these data to predict glaucoma progression and have shown some predictive power Baxter et al. (2019) Banna et al. (2022). However, granular clinical data such as patient presenting symptoms, medical and surgical history, and examination findings are difficult to extract and integrate into predictive models, as these data typically reside within free-text clinical notes in unstructured formats. Recent advances in natural language processing (NLP) have enabled dozens of studies in various medical fields Chen et al. (2020), Mohammadi et al. (2020), ... to get promising results using BERT-based models. However, to our knowledge these techniques have never been investigated for surgery outcome prediction in ophthalmology which reinforced our interest in the topic.

Contribution: We applied BERT-based models and cutting-edge NLP techniques to glaucoma surgery outcome prediction. We went beyond the traditional approaches in ophthalmology (which usually only focus on transformers fine-tuning) by pre-training the network on a set of ophthalmology-related progress and notes and introducing new modeling approaches. We developed various multi-modal fusion models (sum, learnable sum, multiply, concatenate, multi-layer concatenation) leveraging both HER structured data and operative notes. We then used these architectures to develop a model that combine BERT's output into a recurrent neural networks (RNN) and implemented time embeddings (time2Vec) to exploit the temporal nature of the relationship between surgery outcome prediction and physician notes. Our key contribution was to show that these pre-operative progress notes hold some predictive power, a statement that had never been investigated to date. We also confirmed that SOTA methods can not yield outstanding results on tasks where humans are struggling, especially in case like that where inputs are multi-modal.

3 Related Work

NLP Transformers for Ophthalmology and Surgery Outcome Prediction Existing studies on glaucoma surgery prediction only focused on traditional machine learning model (regression, random forest, ...) Baxter et al. (2019) Banna et al. (2022), and have shown decent yet limited predicting power. As off 2022, there is 19 published studies using NLP techniques in ophthalmology Chen and Baxter (2022), including transfer learning with transformers Hu and Wang (2022), supervised approaches leveraging regular expressions Barrows Jr et al. (2000), ... To our knowledge, none of them introduced change in the transformer architecture or took into account time embeddings.

LSTM Models

There are many studies that used LSTMs to classify clinical text data such as Luo (2017). However, from current research we know most SOTA for NLP processing use transformers with the caveat that transformers struggle with large time series. Such as a large corpus of notes that include multiple patient visits over time. A bigger challenge we had was that time series generally tend to struggle with learning these multivariate temporal representations for classification without a large amount of labelled data - this problem inspired us to pre-train an already pre-trained model for our task.

There were several PAPERS that use BERT embeddings and an LSTM in tandem to encode the time representations of other time series data Chaudhury and Sau (2023) Wankhade and Rao (2022). For example, a study found that using a model that generated time embeddings from temporal data and then word embeddings from the notes that were then concatenated and fed into softmax set the new benchmark for predicting in hospital mortality Deznabi et al. (2021). We were particularly inspired by this approach even in the fusion approach where we concatenated different ouptuts from modalities that were then used for classification, we were curious if different types were particularly better. This study also inspired us to think more in-depth about the temporal nature of our data, which led us to trying to separately learn the time embeddings - a simpler but different version of what was done here in time2vec Deznabi et al. (2021).

4 Approach

Task: The goal of the task is to predict the outcome (success or failure) of patients who underwent a glaucoma surgery in Stanford clinics during the 2013-22 period. To do that, we tapped into pre-operative progress notes written by ophthalmologists during patient visits at Stanford hospitals (which describe the patient eye evolution over time) and structured inputs (cf data for more details). Surgical success was defined as a combination of decrease of the eye intraocular pressure (IOP), absence of increase in medication and absence of other surgery/ revision surgery in the years following the operation.

Methods: In the following section, we cover the different architectures that we utilized for this project. The initial plan was the see how BERT and BERT with basic fusion model would perform, and then study sparse attention-heads and gradient methods for interpretability on the notes. However, the surgery prediction task proved to be more difficult that what we initially thought, which is likely why there is not much literature on this task. We ended up experimenting with a whole array of approaches - including but not limited to fusion classification, BERT combined with LSTM models, and time-embeddings.

BERT classifier: We use the BERT architecture based upon transformers using self-attention Devlin et al. (2019). The Bert-uncased, Bio-bert models were used as initial baselines and starting points for which we then pretrained on using masking to improve upon benchmarks Lee et al. (2019). We only did Masked Language Modelling for our pretraining as recent research seems to suggest NSP is may not offer statistically significant performance boosts for additional compute Joshi et al. (2019).

For an intuitive technical understanding of BERT let's denote the input sequence as $X = [x_1, x_2, ..., x_n]$, where x_i is the i-th token in the sequence after it was tokenized by the tokenizer with special tokens such as [CLS] to denote beginnings ends etc. In our case, since we had several notes per patients we would concatenate all the notes per decreasing date time (most recent notes first) and trim the string to the 512 maximum required tokens required by BERT. The output of the encoder can be denoted as $H = [h_1, h_2, ..., h_n]$, where h_i is the hidden embedding of the i-th token in the sequence.

The hidden embedding h_i is produced by passing the input sequence X through the multiple layers of the encoder. Each layer has its own set of parameters that are learned during pre-training. The output of each layer is given by:

$$H^{(l)} = \text{EncoderLayer}(H^{(l-1)})$$

where $H^{(0)} = X$ and $H^{(L)}$ is the output of the final layer (L is the total number of layers). The function EncoderLayer represents a single layer of the encoder, which consists of self-attention and feedforward neural networks.

The self-attention mechanism computes a weighted sum of the input embeddings, where the weights are determined by a soft alignment between each input embedding and all other input embeddings. Mathematically, the self-attention computation for a single layer can be expressed as:

$$\text{SelfAttention}(H^{(l-1)}) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V are linear projections of $H^{(l-1)}$, and d_k is the dimension of the key vectors. The softmax function normalizes the attention weights across all input embeddings.

The output of the self-attention mechanism is then passed through a feedforward neural network, which applies a non-linear transformation to each embedding independently. Mathematically, the feedforward computation for a single layer can be expressed as:

$$\text{FeedForward}(H^{(l-1)}) = \text{ReLU}(H^{(l-1)}W_1 + b_1)W_2 + b_2$$

where W_1, W_2, b_1, and b_2 are learned parameters of the feedforward network, and ReLU is the rectified linear unit activation function.

Finally, the output of the feed-forward network is passed through a layer normalization operation, which normalizes each embedding independently.

By stacking multiple layers of self-attention and feed-forward neural networks with layer normalization, BERT is able to produce a sequence of hidden embeddings of which the one of interest that we use for our purposes is the CLS token, which captures the semantic understanding/acts as a summary sentence for the text that the embedding is produced for.

The CLS token is key for how classification on our data was performed, which we explain on in-depth in the fusion and RNN architecture sections. From our initial experiments explained in our experiment section, it was quickly apparent that BERT and BioBERT were too large/too general for our task. Thus, we used a free corpus of ophthalmology notes to pretrain the BERT from the

BioBERT weight initialization - so that the model could understand shorthand and other semantics specific to ophthalmology notes better.

For example, given an input sequence $X = [x_1, x_2, ..., x_n]$, 20 percent of the input tokens were masked, denoted by the set M, such that $M \subseteq 1, 2, ..., n$. We replace the masked tokens with the "[MASK]" token, denoted by $[MASK]$, to obtain the masked input sequence $\widetilde{X} = [\widetilde{x}_1, \widetilde{x}_2, ..., \widetilde{x}_n]$, where:

$$\widetilde{x}_i = \begin{cases} [MASK] & \text{if } i \in M\ x_i \\ \text{otherwise} \end{cases}$$

The objective of the MLM is to train the model to predict the original input sequence X given the masked input sequence $\widetilde{X}$ and the context of the sequence. To do this, the model is trained to minimize the negative log-likelihood loss of predicting the original tokens given the masked tokens and the context. Mathematically, the MLM loss can be expressed as:

$$\mathcal{L}\text{MLM} = -\sum i \in M \log P(x_i|\widetilde{X}, \Theta)$$

where Θ represents the parameters of the model and $P(x_i|\widetilde{X}, \Theta)$ represents the probability of predicting the original token x_i given the masked input sequence $\widetilde{X}$ and the model parameters Θ.

Fusion architectures: Once we had our baseline and MLM training setup, we built more advanced concatenation models. We came up with two approaches. In the first one, we retrieved BERT last hidden state using either the first CLS token of the sentence or average pooling of all the tokens. We fed that last hidden state to a feed-forward layer to resize the vector and aggregated it with the last hidden state from the structured neural network using several methods. We tried to concatenate the two vectors, multiply them, sum them, sum them with a variable coefficient (i.e. that could give more importance to the output of BERT/of the plain neural network) and to perform SVD on the matrix composed of the two vectors put aside. Since these models were giving were not yielding the accuracy we expected, we tried another to perform "embedded concatenation" where we would concatenate the last hidden state of the plain neural network directly into each transformer block, after the feed-forward layer.

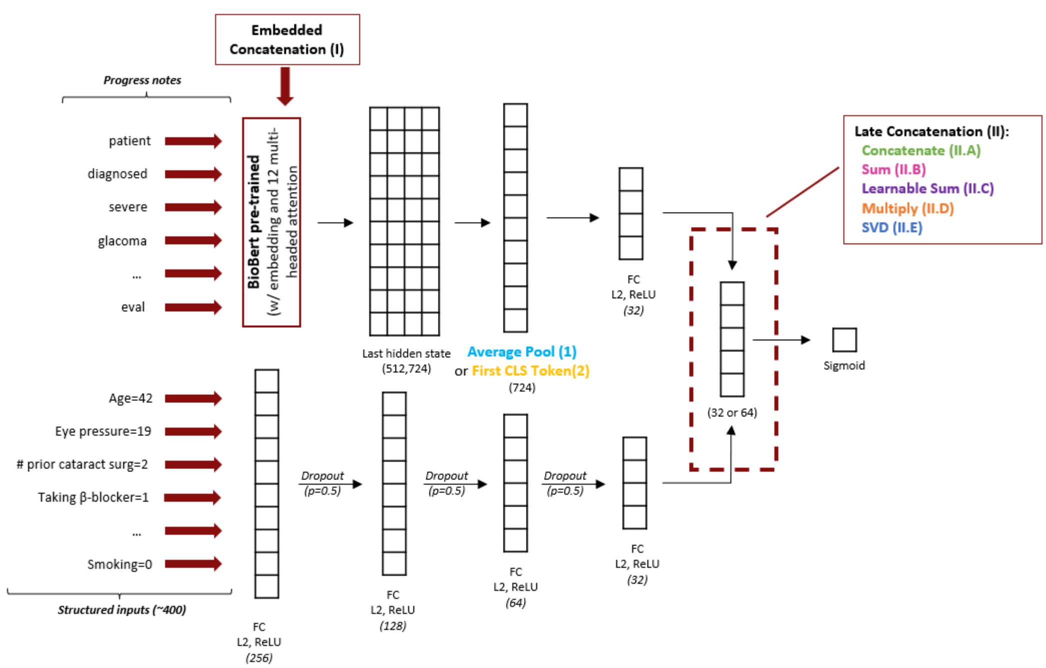

Figure 1: Fusion model architecture

RNN: The goal of the RNN architecture was to capture the potentially temporal nature of the notes' progression for eye disease outcomes, and to avoid the concatenation of all the patient notes into a single one. To do that, pretrained BERT and regular BERT were frozen, and CLS embeddings were generated for each and every note. Then we collated every single embedding for each patient, and created sequences of CLS tokens that correspond to each note over time for each patient. Then these CLS token sequences were classified using a bi-directional LSTM classifier. The pretrained Bio-BERT variant used was frozen so that the gradients does not overfit from learning as that is what occurred when the fusion model above was trained. The primary goal was to see if the LSTM's reccurent capabilities captured the progression in the notes over time better than our pretrained transformer alone, thus we did not pursue extensive optimization of this model due to time constraints.

We take the embeddings that BERT creates and fed them into a bi-directional LSTM. Which then for each token in each sequence of CLS tokens, and sends it through each timestep of the LSTM, from which we extract the last hidden state associated with the last timestep.

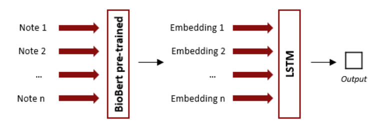

Figure 2: BERT into RNN architecture

Time Embeddings: To improve upon the LSTMs understanding of time, which we later had planned to insert into transformers to replace the learned positional embeddings - we trained a learnable parameter of time via Time2Vec.

For a given scalar notion of time τ , Time2Vec of τ , denoted as $t2v(\tau)$, is a vector of size $k+1$ defined as follows adapted from Kazemi et al. (2019):

$$\text{t2v}(\tau) = \begin{cases} \omega_i \tau + \alpha_i, & \text{if } i = 0 \\ F(\omega_i \tau + \alpha_i), & \text{if } 1 \leq i \leq k \end{cases}$$

Here t2v(τ)[i] is the ith element of t2v(τ), F is a periodic activation function, and ω_i and α_i are learnable parameters.

We added this learnable parameter via a custom layer into the LSTM network. We experimented by multiplying it through the hidden states, and also concacenating it to the input fed into the hidden state. We decided to learn a matrix as an extension simply by changing the number of 1 dimensional features it learned from a series of multi-dimensional features to see if it would increase AUROC - it did not. The choice of sinusoidal activation function F that was used was simply $sin(wj)$, so it could learn the frequency representations of the data.

5 Experiments

5.1 Data

We used a dataset from Stanford medicine data warehouse that contained both structured inputs (e.g. patient age, prior surgeries, ...) as well as free-text operatives notes. We held 20% for test and split the rest between 80% for training and 20% for validation.

The structured data was pulled from various tables and contained 2k rows. It contained 5 types of pre-operative information : prior surgeries, past diagnoses, drug usage, ophthalmology-related clinical measurement and patient general information. One-hot-encoded columns were created (for ophthalmology drugs, general drugs and diagnoses) and variance elimination was performed to only keep the top-100 columns in each category. Clinal measurement (eye intraocular pressure – IOP, best corrected visual acuity – BCVA, . . .) and patient information (weight, height, use of tobacco, . . .) were defined as the last non-null value measured before the operation. As previously mentioned, surgical success was defined as a combination of decrease of the eye intraocular pressure below 19mmHg, decrease (or stagnation) of the number of drug prescribed and absence of other surgery/ revision surgery in the 5 years following the operation.

The unstructured data contained 80k notes, an average of 40 per surgery row. Stop words and punctuation were removed before converting the notes to lowercase.

Xxxx Yyyyy, MD Clinical Assistant Professor of Ophthalmology Cornea & External Diseases Byers Eye Institute at Stanford University 2452 Watson Court Palo Alto, CA 94025 (650) 723-6995 Eye Institute at Stanford University Hospital and Clinics Ophthalmology Outpatient Encounter Diagnoses 1. Secondary angle-closure glaucoma, right, severe stage 2. Iris replaced by transplant 3. Blepharitis of upper and lower eyelids of both eyes, unspecified type Assessment Corneal perforation status post tectonic penetrating keratoplasty (6/5/13), amniotic membrane graft, and lysis of adhesions left eye (6/3/13), perforation due to severe blepharitis/MGD. Peripheral cornea just nasal to nasal graft-host junction with stable thinning and resolved stromal neovascularization. Patient using warm compresses 3/27/2018. Stable graft, no rejection Secondary angle closure glaucoma. Severe vision loss right eye. Status post express shunt with bleb revision left eye and BGI (12/15/14). Intraocular pressure good. Using lotemax once a day only.

Figure 3: Clinal note example (anonymized and masked)

5.2 Evaluation method

As the task consisted in classifying surgery records, we used the typical classification metrics. We reported accuracy, F1 score, precision and recall but ran optimization based on accuracy.

We benchmarked different loss functions (cross entropy, binary cross entropy, mean square error and KL divergence) and cross entropy yielded the best results. We weighted the loss function to account for the class imbalance (60% False vs. 40% True).

5.3 Experimental details

We ran our experiments on 3 different pre-trained models: Bert Base Uncased (a general language model), BioBert (a language model fine-tuned on medical data) and TinyBert (a general language model 7 times smaller than traditional BERT). We ran grid-search on hyper-parameters (learning rate, weight decay and optimizer) and launched experiments with the best-performing value. Learning rate was set to 3×10^{-5}, weight decay to 10^{-2} and the optimizer to Adam with 2 steps of gradient accumulation. We trained all the models on 5 epochs but used early-stopping

To include the structured inputs, we then performed grid-search to select the best feed forward neural network architecture among the 20 tested, and picked a 4-layer one (256-128-64-32) and used ReLU.

Since our result contained a great amount of variability, we performed 5 fold cross-validation and give all of our results averaged over these runs.

5.4 Results

We ran the experiments with the parameters aforementioned for all of our models, and obtained the following results:

Table 1: Benchmark of the Accuracy, AUROC and F1 score of the model benchmarked

Model	**Accuracy**	**AUROC**	**F1**
BERT (baseline)	0.620	0.534	0.211
BERT Pretrained	0.644	0.594	**0.348**
Best Fusion Model	**0.647**	**0.629**	0.312
BERT and LSTM w/o time embedding	0.618	0.518	0.232
BERT and LSTM w/ time embedding	0.616	0.510	0.263

Where the best fusion model is the late concatenation method with first CLS token retrieval. The results of the other fusion models are below.

Overall, we were disappointed with our results. The accuracy we achieve is in line yet slightly below the top-performing paper in the literature which managed to get an accuracy of 0.68 on a similar but more restrained and somewhat simpler task (only looking at trabeculectomy surgeries). However, this paper only used basic machine learning models (regression, random forest, KNN, ...) whereas we deployed a range of more advanced deep learning models. We believe we were not able to outperform this paper because of the difference in the nature of the task, the fact that we used distinct data sets and possibly because the notes did not contain as much information as we expected.

Table 2: Deep dive on fusion model results

Model	Accuracy	AUROC	F1
Retrieval Approach: Average Pooling	0.624	0.598	0.278
Retrieval Approach: First CLS Token	**0.633**	**0.614**	**0.312**
Embedded Concatenation	0.618	0.523	0.212
Late Concatenate	**0.647**	**0.629**	0.312
Late Sum	0.628	0.624	0.303
Late Learnable Sum	0.631	0.611	0.309
Late Multiply	0.636	0.596	**0.368**
Late SVD	0.636	0.610	0.302

In terms of performance difference between our models, we were surprised to see that only one fusion approach manages to increase accuracy vs. the plain BERT pre-trained model, with approach such as embedded concatenation scoring even worse than our baseline. Similarly, feeding the last hidden state of BERT in a LSTM seemed like a great idea but performed worse vs. our baseline and ended up almost being a random guesser since with a ROC score of 0.51. In our case, "less was more" since the plain pre-trained model already greatly increased the accuracy of the plain BERT model, and the other approaches were only able to enhance it a little bit.

Finally, it is worth mentioning that our results were very noisy - and especially the F1 score since the majority class was False and the classifier rarely predicted positive results. We averaged all of our results over 5 cross validation runs to get a more stable picture of the model performances, but the underlying issue of model variability remains present.

6 Analysis

Despite all our experimentation, our models did not meet the accuracy and AUROC score we were expecting. We thus believe that their is not enough predictive information in the notes to make great predictions with regard to the size of the dataset and the complexity of the task. We could potentially improved the model accuracy if we could feed him with more data (since fusion and pre-trained have some predictive power), yet it is unlikely to have much bigger dataset for such a specific task.

Taking a deeper look at the data, we noticed a pattern along the notes - a lot of them had a great amount of content copied and pasted from the patient previous note (to keep the history of the patient encounters in the note) with only the few last lines being new. This is not a problem unique to our dataset, and we realized it was a common clinical practice Steinkamp et al. (2022). Our dataset could easily be 50% smaller, as recent studies show that 50 % of all clinical notes are simply duplicates of information. Thus, our hypothesis about time embeddings might not have panned out for this particular task, as there might not have been strong markers of prediction in all the notes over time - and even if there were - they were not recorded in enough notes as a great amount are duplicated.

7 Conclusion

Based on our experiments, the best performing model is the late fusion concatenation one which leverages the structured data. This model not only had a slightly better accuracy vs text-only approaches, but also a smallest variance in terms of accuracy and AUROC. We believe that the additional structured information is slightly increasing and somehow stabilizing the predictive power of the BERT-type model which contains a lot more noise.

Our work outlined the basic approaches and some of the problem that arise when using NLP for this particular task and data - as this is the first study in the literature to handle this task with deep learning to our knowledge. Even if we did not reached the level of accuracy expected, we showed that NLP models can extract some predictive power from the free-text notes. Finally, although no paper explicitly explored that, our supervisor Dr. Wang insisted on the fact that human performance is typically "very very poor" on this type of task. Extracting some information from these notes is thus already an encouraging result.

A promising next step could be to train larger multi-modal systems with notes, text, and potentially image data on all sorts of medical problems. We could then potentially use them via prompting to come to inferences, and perform some few shot learning via prompting. We believe this is an interesting path since our work clearly demonstrated that more general models like transformers which learn positional embeddings outperform models that methods that are more specific such as Time2Vec to learn time embeddings, and even LSTMs.

Finally, as pretraining is clearly the process that provided the model with a lot of knowledge, it could also be interesting to focus on the data to retrieve larger or more specific database to further train our existing models.

References

Hafiz U Banna, Ahmad Zanabli, Benjamin McMillan, Ruslan Kamalov, Shariq Ali, Rahul Kukreja, Maninder Singh, Tehmina Arif, Kemal Ornek, Carlos Gustavo De Moraes, et al. 2022. Evaluation of machine learning algorithms for trabeculectomy outcome prediction in patients with glaucoma. *Scientific reports*, 12(1):1–10.

Randolph C Barrows Jr, Mihai Busuioc, and Carol Friedman. 2000. Limited parsing of notational text visit notes: ad-hoc vs. nlp approaches. In *Proceedings of the AMIA Symposium*, volume 2000, pages 51–55. American Medical Informatics Association.

Stephanie L Baxter, Caroline Marks, Tina-Tl Kuo, Lucila Ohno-Machado, and Robert N Weinreb. 2019. Machine learning-based predictive modeling of surgical intervention in glaucoma using systemic data from electronic health records. *American Journal of Ophthalmology*, 208:30–40.

Sushovan Chaudhury and Kartik Sau. 2023. A bert encoding with recurrent neural network and long-short term memory for breast cancer image classification. *Decision Analytics Journal*, 6:100177.

Jennifer S Chen and Seth L Baxter. 2022. Applications of natural language processing in ophthalmology: Present and future. *Frontiers in Medicine*, 9:906554.

Lijun Chen, Raj Shah, Thomas Link, Matthew Bucknor, Sharmila Majumdar, and Valentina Pedoia. 2020. Bert model fine-tuning for text classification in knee oa radiology reports. *Osteoarthritis and Cartilage*, 28(S1):S315–S316.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Iman Deznabi, Mohit Iyyer, and Madalina Fiterau. 2021. Predicting in-hospital mortality by combining clinical notes with time-series data. In *Findings*.

Wendeng Hu and Sophia Y Wang. 2022. Predicting glaucoma progression requiring surgery using clinical free-text notes and transfer learning with transformers. *Translational Vision Science Technology*, 11(3):37.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2019. Spanbert: Improving pre-training by representing and predicting spans. *CoRR*, abs/1907.10529.

Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus A. Brubaker. 2019. Time2vec: Learning a vector representation of time. *CoRR*, abs/1907.05321.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *CoRR*, abs/1901.08746.

Yuan Luo. 2017. Recurrent neural networks for classifying relations in clinical notes. *Journal of Biomedical Informatics*, 72:85–95.

Rahim Mohammadi, Shashank Jain, Ali Tavakoli Namin, Farzaneh Karami, Neil S Kamdar, Cody Foltz, James Huddleston, Suchi Saria, and Lillian Kao. 2020. Predicting unplanned readmissions following a hip or knee arthroplasty: Retrospective observational study. *JMIR medical informatics*, 8(8):e19761.

Jackson Steinkamp, Jacob J. Kantrowitz, and Subha Airan-Javia. 2022. Prevalence and Sources of Duplicate Information in the Electronic Medical Record. *JAMA Network Open*, 5(9):e2233348–e2233348.

Yih-Chung Tham, Xin Li, Tien Yin Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. 2014. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11).

Mayur Wankhade and Annavarapu Rao. 2022. Opinion analysis and aspect understanding during covid-19 pandemic using bert-bi-lstm ensemble method. *Scientific Reports*, 12.