Author/s:
Tahayori, B;Chini-Foroush, N;Akhlaghi, H

Title:
Advanced natural language processing technique to predict patient disposition based on emergency triage notes

Date:
2021-06-01

Citation:
Tahayori, B., Chini-Foroush, N. & Akhlaghi, H. (2021). Advanced natural language processing technique to predict patient disposition based on emergency triage notes. EMA Emergency Medicine Australasia, 33 (3), pp.480-484. https://doi.org/10.1111/1742-6723.13656.

Persistent Link:
https://hdl.handle.net/11343/276429

Tahayori Bahman (Orcid ID: 0000-0002-4927-0023)
Akhlaghi Hamed (Orcid ID: 0000-0002-3929-7265)

Title Page

Manuscript category: Original Research

Title: An advanced natural language processing technique to predict patient disposition based on emergency triage notes

Short running title: Predicting patient disposition

Key words: artificial intelligence, natural language processing, patient disposition, triage notes

Authors:

Bahman Tahayori: PhD in Biomedical Engineering, Department of Biomedical Engineering, The University of Melbourne, and St Vincent's Hospital, Fitzroy, Vic, Australia, Email: bahmant@unimelb.edu.au

Noushin Chini-Foroush: Medical Doctor, Department of Neurology, Monash Health, Clayton, Vic, Australia, Email: noushin.chiniforoush@gmail.com

Hamed Akhlaghi: Director of Emergency Research, Emergency physician, St Vincent's Hospital, Fitzroy, Vic, Australia, Email: hamed.akhlaghi@svha.org.au

Corresponding author:
Bahman Tahayori, Department of Biomedical Engineering, The University of Melbourne
Address: Room 1.29, KMB, The University of Melbourne, Parkville, Vic, 3010, Australia
Email address: bahmant@unimelb.edu.au

Authors' contribution:
B Tahayori contributed in developing the machine learning model, python programming as well as writing the paper. N Chini-Foroush contributed in refining the manuscript. H Akhlaghi proposed the concept and contributed in writing the paper and designing the model.

Ethics: This study was approved by the ethics committee at St Vincent's Hospital (Melbourne) Limited with a local reference number 240/19.

Acknowledgment: This project received a research endowment fund from St Vincent's research department.

Competing interests: The authors declare no competing interests.

Data availability: Data available on request from the authors.

Emergency Medicine Australasia

ORIGINAL RESEARCH

An advanced natural language processing technique to predict patient disposition based on emergency triage notes

Bahman Tahayori[1,2], Noushin Chini-Foroush[3], Hamed Akhlaghi[2]

1) Department of Biomedical Engineering, The University of Melbourne, Parkville, Vic, Australia

2) Emergency Department, St Vincent's Hospital, Fitzroy, Vic, Australia

3) Department of Neurology, Monash Health, Clayton, Vic, Australia

Email addresses:

Bahman Tahayori: bahmant@unimelb.edu.au,

Noushin Chini-Foroush: noushin.chiniforoush@gmail.com,

Hamed Akhlaghi: hamed.akhlaghi@svha.org.au

Abstract

Objective: To demonstrate the potential of machine learning and capability of natural language processing to predict disposition of patients based on triage notes in the emergency department (ED).

Methods: A retrospective cohort of ED triage notes from St. Vincent's Hospital (Melbourne) was used to develop a deep-learning algorithm that predicts patient disposition. Bidirectional Encoder Representations from Transformers (BERT), a recent language representation model developed by Google, was utilised for natural language processing. 80% of the dataset was used for training the model and 20% was used to test the algorithm performance. ktrain library, a wrapper for TensorFlow Keras, was employed to develop the model.

Results: The accuracy of the algorithm was 83% and the Area Under the Curve (AUC) was 0.88. Sensitivity, specificity, precision and F1-score of the algorithm were, 72%, 86%, 56% and 63%, respectively.

Conclusions: Machine-learning and natural language processing can be together applied to the ED triage note to predict patient disposition with a high level of accuracy. The algorithm can potentially assist ED clinicians in early identification of patients requiring admission by mitigating the cognitive load, thus optimises resource allocation in EDs.

Key words: triage notes, patient disposition, natural language processing, artificial intelligence

Ethics: This study was approved by the ethics committee at St Vincent's Hospital (Melbourne) Limited with a local reference number 240/19.

Acknowledgment: This project received a research endowment fund from St Vincent's research department.

An advanced natural language processing technique to predict patient disposition based on emergency triage notes

1. Introduction

Patient presentations to Emergency Departments (EDs) have gradually increased nationally and internationally along with the aging population.[1-3] Early decision-making can potentially improve the management of ED flow through the effective delivery of high standard, quality care to patients and reducing the length of stay in the ED.[4, 5]

To optimise service provision in EDs, several innovative approaches have been trialled.[6] Applying artificial intelligence (AI) to the medical field, particularly to ED data, is a novel area of research and has shown promising results.[7-9] Natural Language Processing (NLP) is one of the main components of AI which converts unstructured data, such as ED triage notes, to a set of quantitative parameters.[10]

Taken by a trained emergency nurse, an ED triage note contains the narrative of a patient's presenting illness or a reason to present to the ED. Several studies have recently applied Artificial Neural Networks (ANN) and its variants together with NLP algorithms to triage notes to predict patient disposition.[11-13] The results are promising, with up to three quarters of the outcomes being predicted correctly depending on the NLP technique utilised.[12]

Dinh et al. used the Sydney Triage to Admission Risk Tool (START) to predict inpatient admission using the patient's age and presenting illness as triage categories. [14] They achieved an accuracy of 75.24% using regression analysis. Sterling et al. have used NLP to predict patient disposition based on triage notes.[12] They reported the accuracy of 68%, 73% and 62% for three different methods, Bag of Words (BOW), Paragraph Vectors (PV) and Topic Models (TM), respectively. Similar approaches have been applied to subsets of ED patients in more specialised areas, such as the paediatric population or asthmatic patients.[15] These studies show the feasibility of implementing AI in EDs to predict the final discharge planning for patients at the time of presentation. More recently, Fernandes et al. applied machine learning and NLP techniques to triage data and classified patients with high mortality and cardiopulmonary arrest.[16] Bacchi et al. employed machine-learning and NLP algorithms to predict patients' length of stay using triage notes with 82% accuracy. [17]

In this study, we utilised AI and NLP to predict inpatient admission using unstructured ED triage notes at St Vincent's Emergency Department. To our knowledge, this is the first Australian study implementing **the Bidirectional Encoder Representations from Transformers (BERT), a novel state-of-the-art AI model that was shown to improve many NLP tasks[18, 19],** to train an AI system to analyse the free-text ED triage notes and predict patient disposition.

2. Methods

2.1. Data: A retrospective cohort of consecutive ED presentations to St. Vincent's Hospital (Melbourne) from July 2010 to June 2019 were used. We excluded data where the triage note was blank or the disposition was not "Home" or "Admit to Ward". **Particularly, we did not consider Emergency Short Stay (ESS) admissions in our model.**

2.2. Natural Language Processing: We used NLP to convert triage note data to numerical vectors that can subsequently be used by machine-learning algorithms to predict patient disposition.

2.3. Pre-processing: A crucial step in machine learning is data cleaning and pre-processing. We only kept "Triage Note" and "Disposition" columns from the dataset as the input for the machine learning algorithm. Prior to employing NLP techniques, we stemmed and lowercased data, removed the punctuation and delete the stop words **(common words that are excluded prior to applying NLP, e.g. "the", "that" and "is")**.

2.4. The BERT model: We used BERT, a recent language representation model developed by Google.[18, 19] To implement and train the deep-learning algorithm, we used ktrain library which is a wrapper for TensorFlow Keras.[20]

2.5. Machine learning steps: We used 80% of the dataset as the training set. To minimise the correlation, the split of the dataset was performed on a random basis. The training dataset was used to train the model and tune its parameters. Given that the dataset is imbalanced in terms of the two disposition classes, we balanced the dataset by oversampling the minority class "Admit" to avoid naive results with high accuracy. The block diagram of the modelling process is illustrated in Figure 1. We used the "fit_onecycle" method in ktrain to train our model. We set the "1 cycle learning rate" to 2×10^{-5}. The parameters for training the model are summarised in Table 1.

The remaining 20% of the dataset was used as the test set to evaluate the performance of the model for unseen observations. The test dataset was not balanced because it should reflect the reality of the data.

2.6. Evaluation by experts: We used input from ED consultants to find a reference for comparing the algorithm performance with experts. Two hundred triage notes with only "Home" or "Admit" dispositions were randomly sampled from the original dataset and distributed among five ED consultants for them to manually assign disposition. The unlabelled datasets only contained the real data without any pre-processing. **The overall performance of the experts was calculated by forming a confusion matrix for each expert separately and averaging the results.**

3. Results

There were 399,876 consecutive ED presentations between July 2000 to June 2019. After excluding blank triage notes and dispositions other than "Home" and "Admit to Ward", a total of 249,532 ED encounters were used for the AI analysis. The "Home" and "Admit to Ward" dispositions have 203,693 (81%) and 45,839 (19%) records respectively. There is no significant difference between the length of the sentences in the triage notes of the two groups. The median number of words are 26 and 28 for "Home" and "Admit to Ward" respectively. The maximum length of the triage notes was 60 words. The accuracy of the model was 83% and the Area Under the Curve (AUC) was 0.88. The Receiver Operating Characteristic (ROC), which shows the True Positive Rate (TPR) versus the False Positive Rate (FPR), is demonstrated in Figure 2. Assuming that "Admit" is positive and "Home" is negative, the detailed performance of the model, including recall (sensitivity or true positive rate), specificity (true negative rate), precision (positive predictive value), F1-score (the harmonic mean of precision and sensitivity) and accuracy of both the AI algorithm and the human experts is presented in Table 2. In predicting the disposition only by reading the triage note without any access to the patient demographics or vital signs, the accuracy of the ED consultant was 78% with a specificity of 77%. Results for the performance of each emergency physician are shown in Table 3.

4. Discussion

In this study, we designed and implemented an optimal AI-based algorithm to predict patient disposition with a level of accuracy comparable to that of experienced emergency physicians. The top-four frequent words in both the groups ("Home" and "Admit") were similar (see Appendix I). This simple analysis confirms that NLP employs more advanced techniques and statistical analyses to capture the content and meaning of triage notes to predict disposition. The designed AI algorithm is capable of extracting the context of each triage note to determine the disposition. We employed deep learning and the BERT model to predict patient disposition solely based on triage notes.

Medical electronic systems hugely vary between hospitals in different states of Australia. Some sophisticated systems allow the user, a trained ED nurse in this situation, to input the patient's vital signs or other relevant information enclosed with an ED triage note. But older systems allow adding only free-text triage notes in which the ED nurse can write the patient's vital signs and past medical history as free text. Therefore, we decided to purposefully train our algorithm to use only free-text triage note as an input without considering any other influential parameters such as age and vital signs. This allows the algorithm to be easily and readily added to most of the ED electronic records without the need to alter the currently installed system. The reasons for the improved performance of the AI predictor metrics presented in this paper compared to Sterling's study[12] can be explained by several factors:

1) The difference in the datasets generated by triage nurses with different training; for example, the average length of triage notes reported in Sterling's work was 64.3 words (SD = 35.2) while it was 27 words in our dataset (SD = 6.9).
2) The electronic patient record at St Vincent's Hospital does not allocate a column to add patients' vital signs separately. Therefore, this information is written as a free text in the triage note.
3) While we used data from one centre, Sterling's study used data from three different centres. This could potentially result in more complications due to differences in triage nurse training and the culture of each centre.
4) We applied the BERT model to triage data. Given that BERT significantly outperforms other available methods,[21] we believe that applying BERT is a key factor in predicting patient disposition with higher performance metrics. To verify this, we applied BOW to our data and it resulted in an accuracy of 72% and an AUC of 0.77, which is similar to the metrics reported in Sterling's work.

The result presented in Table 3 indicates that the AI algorithm is more specific than human experts. On the other hand, human experts are more sensitive than the AI algorithm, meaning that their preference is to not miss the "Admit" dispositions. The algorithm is more specific and less sensitive because the test dataset is imbalanced, and the size of the "Admit" class is approximately a quarter of the "Home" class.

F1 score is a measure that determines how well a classifier distinguishes between two classes and is the mean of sensitivity and precision rates. In terms of the F1-score, the AI algorithm's performance is very close to that of the experts. We expect that adding age as a separate distinctive feature in predicting patient disposition using multi-parameter classification techniques will further improve the performance of the designed AI algorithm **as well as the**

expert performance. The patient history, if available, can be considered in AI models to further enhance the algorithm performance.

Theoretically, the system should be able to improve the National Emergency Access Target (NEAT).[5] Early identification of patients requiring admission will also optimise resource allocation by streamlining the care and preventing double handling. Furthermore, if a patient is flagged as "Admit" on arrival to the ED based on this algorithm, the ED nurse in-charge (NIC) can expedite ED bed allocation for early review. Therefore, bed allocation begins at the time of presentation, providing the hospital administrators sufficient time to secure a bed.

A potential advantage of the system following improving its performance by considering age and patients' history is that it can act as a safeguard against premature discharge. If a patient is discharged from the hospital within 24 hours of presentation despite the AI system flagging the case as "Admit", the discharge planning should be reviewed by a senior clinician to ensure the safety of the patient and appropriateness of the plan. This process will subsequently promote quality care. Discharge planning is a complex process requiring inputs from a multi-disciplinary team consisting of physicians, pharmacists, physiotherapists and allied health clinicians. Unplanned re-admission costs around 20 billion dollars annually in the US.[22] The figures are not better in Australia with the cost going up to 431 million dollars for the state of Victoria and 1.5 billion dollars across Australia. Re-admission and utilisation of health care services, such as nurse or general practitioner visits to facilities, renders the cost-saving measures of premature discharges ineffective.[23]

We excluded Emergency Short Stay (ESS) admissions from the database because the model of care differs amongst hospitals. In addition, our hospital utilises the ESS for both admitted and non-admitted patients; therefore, making a distinction between these two groups is not possible for the AI. Further analysis is required to classify ESS admission to either “Admit” or “Home” disposition, and then the data can be fed to the AI algorithm.

5. Limitations

This is a single study design to predict admission based on the triage notes. St Vincent’s Hospital has an acute adult ED with no obstetric and paediatric presentations. Further multi-centre study is required to evaluate the precision of the designed model in other health care services. We suspect that, due to differences in the way the triage note is written in different EDs, the AI system has to be re-trained using the local data to improve its accuracy before implementation. Besides, it is highly plausible that the performance of the system can improve by the use of a larger dataset from a multi-centre study. Given that the information of the patient’s vital signs is written as a free-text in the triage note, the system may identify a pattern between the vital signs and the disposition. An evaluation of the current trained system using data from other health services where vital signs are not incorporated in the triage note may be able to clarify this.

6. Conclusion and future work

In this paper, we showed that machine-learning and NLP techniques can be together applied to the triage note data generated by trained triage nurses at EDs to predict patient disposition **with an acceptable level of accuracy. Further tuning of the algorithm is required to achieve higher performance.** We propose conducting a multi-centre (including hospitals with paediatric presentations and obstetric emergencies) AI analysis to further evaluate the performance and reliability of the proposed AI system. Age as well as patient history data can be incorporated to further improve the outcome of the model. To use AI algorithms clinically, the ethics of using digital tools in health care systems should be taken into account.[24]

7. Acknowledgment

This project received a research endowment fund from St Vincent's research department.

References

1. Gillam S. Rising hospital admissions. BMJ. 2010;340:c636.
2. Australian Institute of Health Welfare, Emergency department care 2017–18: Australian hospital statistics. Canberra: AIHW; 2018.
3. Mason S, Mountain G, Turner J, Arain M, Revue E, Weber EJ. Innovations to reduce demand and crowding in emergency care; a review study. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine. 2014;22(1):55.
4. Ebker-White A, Bein KJ, Berendsen Russell S, Dinh MM. The Sydney triage to admission risk tool (START) to improve patient flow in an emergency department: a model of care implementation pilot study. BMC Emergency Medicine. 2019;19(1):79.
5. Rendell K, Koprinska I, Kyme A, Ebker-White AA, Dinh MM. The Sydney Triage to Admission Risk Tool (START2) using machine learning techniques to support disposition decision-making. Emergency Medicine Australasia. 2019;31(3):429-35.
6. Barak-Corren Y, Israelit S, Reis B. Progressive prediction of hospitalisation in the emergency department: Uncovering hidden patterns to improve patient flow. Emergency Medicine Journal. 2017;34:emermed-2014.
7. Faita F. Deep learning in Emergency Medicine: Recent contributions and methodological challenges. Emergency Care Journal. 2020;16.
8. Roquette BP, Nagano H, Marujo EC, Maiorano AC. Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. Neural Networks. 2020;126:170-7.
9. Wang F, Preininger A. AI in Health: State of the Art, Challenges, and Future Directions. Yearb Med Inform. 2019;28(1):16-26.
10. Eisenstein J. Introduction to Natural Language Processing: MIT Press; 2019.
11. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Critical Care. 2019;23(1):64.
12. Sterling NW, Patzer RE, Di M, Schrager JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. International Journal of Medical Informatics. 2019;129:184-8.
13. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of Emergency Department Hospital Admission Based on Natural Language Processing and Neural Networks. Methods Inf Med. 2017;56(5):377-89.
14. Dinh MM, Russell SB, Bein KJ, Rogers K, Muscatello D, Paoloni R, et al. The Sydney Triage to Admission Risk Tool (START) to predict Emergency Department Disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia. BMC Emergency Medicine. 2016;16(1):46.
15. Goto T, Camargo CA, Jr., Faridi MK, Freishtat RJ, Hasegawa K. Machine Learning–Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. JAMA Network Open. 2019;2(1):e186937-e.

16. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. PLOS ONE. 2020;15(4):e0230876.
17. Bacchi S, Gluck S, Tan Y, Chim I, Cheng J, Gilbert T, et al. Prediction of general medical admission length of stay with natural language processing and deep learning: a pilot study. Internal and Emergency Medicine. 2020.
18. Devlin J, Chang M, Toutanova K, editors. 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; 2019.
19. Li F, Jin Y, Liu W, Rawat BPS, Cai P, Yu H. Fine-Tuning Bidirectional Encoder Representations From Transformers (BERT) Based Models on Large-Scale Electronic Health Record Notes: An Empirical Study. JMIR Medical Informatics 2019;7.
20. Maiya AS. ktrain: A Low-Code Library for Augmented Machine Learning. arXiv e-prints [Internet]. 2020 April 01, 2020:[arXiv:2004.10703 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2020arXiv200410703M.
21. Young T, Hazarika D, Poria S, Cambria E. Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. IEEE Computational Intelligence Magazine. 2018;13:55-75.
22. Jencks S, Williams M, Coleman E. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. The New England journal of medicine. 2009;360:1418-28.
23. Considine J, Fox K, Plunkett D, Mecner M, O'Reilly M, Darzins P. Factors associated with unplanned readmissions in a major Australian health service. Australian Health Review. 2017;43.
24. Burr C, Taddeo M, Floridi LJS, ethics e. The Ethics of Digital Well-Being: A Thematic Review. Scienece and Engineering Ethics. 2020.

Appendix I: Insight into data

We depicted a word cloud for triage notes with “Home” and “Admit to Ward” dispositions in Figures A1(a) and A1(b), respectively. Although there are some differences between the word clouds of these two outcomes, there is no clear distinction between the two clouds. This shows that merely keywords are not accurate indicators of patient disposition; therefore, advanced natural language processing methods are required to make more accurate predictive models.

Table A1 shows the top five words/phrases in the word cloud for “Home” and “Admit to Ward” classes separately. There is no significant difference between the number of words in the triage notes with different dispositions. The median numbers of words for triage notes with “Home” and “Admit to Ward” dispositions are 26 and 28 words, respectively. The maximum number of words used in a triage note is 60 words for this dataset. As with keywords, the number of words is not indicative of patient disposition; thus, advanced tools are required to predict disposition.

(a)

(b)

Figure A1- Word cloud of the triage note data for (a) triage notes with “Home” disposition and (b) triage notes with “Admit to Ward” disposition.

Table A1: The 5 words/phrases with maximum frequency in triage notes with “Home” and “Admit to Ward” dispositions.

Home	**Admit**
Abdo pain	Abdo pain
Today	Today
Last night	Biba
Chest pain	Pain
Back pain	Post

Table 1: The parameters used to train the model. “Maximum features” specifies the maximum number of different words in the dataset. The maximum number of words considered for each triage note is determined by “maximum length”. Number of training samples used in each iteration is indicated by “batch size”.

Model parameter	**Value**
Maximum features	50000
Maximum length	64
Classes	“Home” and “admit”
Batch size	12
1 cycle learning rate	2×10^{-5}

Table 2: Performance of the AI algorithm compared to expert results.

	AI Algorithm	Human Experts
Accuracy	0.83	0.78
Specificity	0.86	0.77
Precision	0.56	0.5
Recall	0.72	0.9
F1-score	0.63	0.64

Table 3: Performance of human experts (E1-E5).

	E1	E2	E3	E4	E5
Accuracy	0.85	0.75	0.75	0.77	0.8
Specificity	0.87	0.69	0.75	0.75	0.78

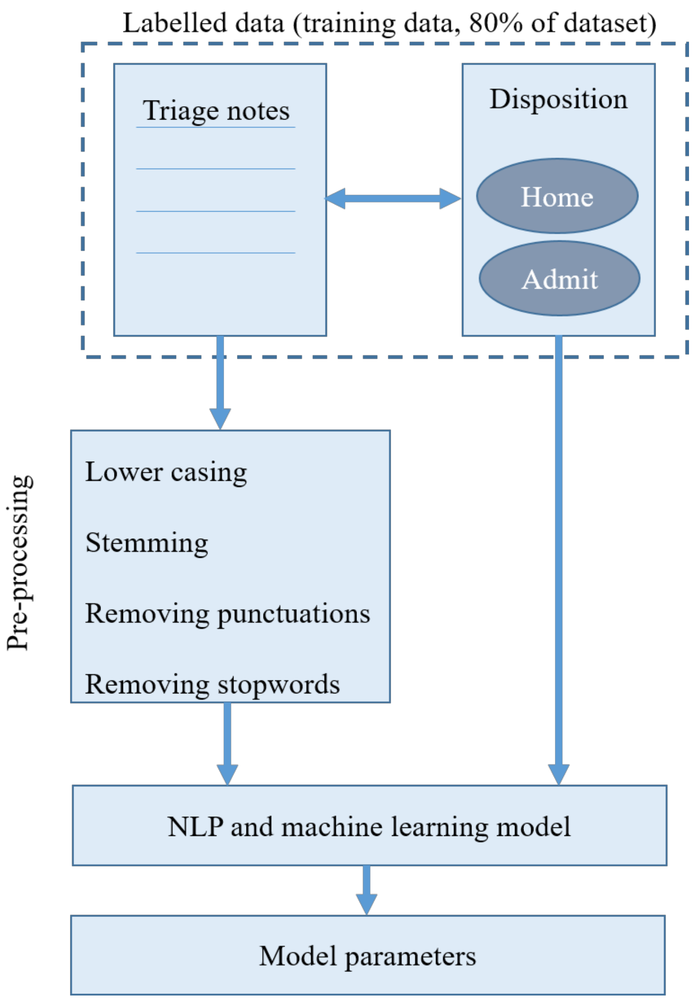

Figure 1- The block diagram of the machine learning algorithm showing the steps of the process.

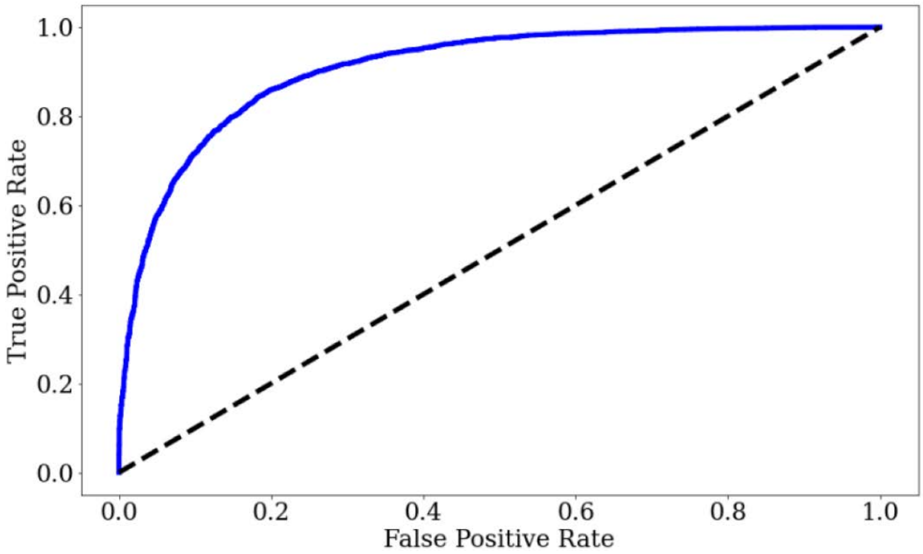

Figure 2- The ROC curve of the trained model on the test dataset that shows the performance of the classifier. The AUC of this curve is 0.88.