AI-Enabled Automation Solution for Utilization Management in Healthcare Insurance

Gaurav Karki[1[0000-0003-0445-3261]], Jay Bharateesh Simha[2[0000-0002-6733-7535]],

Rashmi Agarwal[3[0000-0003-1778-7519]]

[1,2,3] REVA Academy for Corporate Excellence (RACE), REVA University
Bengaluru, India
[1]gauravk.ai01@race.reva.edu.in,
[2]jb.simha@reva.edu.in,
[3]rashmi.agarwal@reva.edu.in

Abstract. As businesses advance toward digitalization by automating an increasing number of procedures, unstructured forms of text in documents present new challenges. Most organizational data is unstructured, and this phenomenon is on the rise. Businesses like healthcare and insurance are embracing business process automation and making considerable progress along the entire value chain. Artificial intelligence (AI) algorithms that help in decision-making, connect information, interpret data, and apply the insights gained to rethink how to make better judgments are necessary for business process automation.

A healthcare procedure called Prior Authorization (PA) could be made better with the help of AI. PA is an essential administrative process that is a component of their utilization management systems, and as a condition of coverage, insurers require providers to obtain preapproval for the provision of a service or prescription. The processing of insurance claim documents can be facilitated using Natural Language Processing (NLP). This paper describes the migration of manual procedures to AI-based solutions in order to accelerate the process. The use of text similarity in systems for information retrieval, question-answering, and other purposes has attracted significant research. This paper suggests using a universal sentence encoder, a more focused strategy, to handle health insurance claims. By extracting text features, including semantic analysis with sentence embedding, the context of the document may be determined. The outcome would have a variety of possible advantages for members, providers, and insurers. AI models for the PA process are seen as promising due to their accuracy and speed of execution.

Keywords: Utilization Management, Prior Authorization, Healthcare, Insurance, Claim Processing, Deep Learning, Artificial Intelligence, Automation, Natural language processing.

1 INTRODUCTION

The expenses of the healthcare system have been spiraling out of control for years, and utilization management is a crucial method for insurers and providers to guarantee that adequate care is delivered cost-effectively. Utilization management (UM) is the evaluation of medical care based on evidence-based criteria and health payer requirements. In addition to reducing costs, the goal is to give patients the appropriate treatment, from a shorter duration of stay to enhanced release planning. Once insurers have defined their norms and guidelines, the utilization management process centers on controlling prior authorization filings via clinical and peer-to-peer evaluations [1].

The influence of UM processes on payer finances, case management, and health plan member and provider satisfaction is direct. Ineffective utilization control results in annoying care delays and higher operating expenses for health insurance. However, conventional UM systems place member and provider satisfaction and cost control in opposition. More rules or more comprehensive review processes can reduce the high cost of medical care, but at the expense of member and provider satisfaction. Up until now, there has always been a compromise. Now, insurers have the chance to transform their UM processes through the application of automation and AI. By introducing an AI-based solution for utilization management, insurers can simultaneously enhance the member and provider experience while reducing operating expenses and medical costs [2].

PA is a primary process of UM. This is conducted before to at the start of treatment on a particular scenario basis to eliminate needless services. The selected treatment should be considered provisional and subject to modification in the future. PA is an examination of a patient's condition and suggested therapy. Its primary objective is to reduce unnecessary, ineffective, or redundant treatments. PA is utilized for regular and urgent referrals, but not for emergency room admissions. The review might take place either before or after admission, but always before treatment begins. In some situations, a physician's directions may not be followed, which could enrage both the medical staff and the patient.

This paper proposed solution tries to automate the PA procedure. Unlike basic Robotic Process Automation (RPA) systems that just automate individual aspects of the prior authorization process, artificial intelligence may fundamentally alter the way reviewers handle prior authorization requests. If a healthcare provider submits an authorization request, artificial intelligence evaluates this information to the medical necessity criteria to ensure that the patient receives the proper care. If all conditions are satisfied, the request may be automatically authorized, without the requirement for a utilization management reviewer to touch the previous authorization and a clinical evaluation. This can reduce approval times from weeks to hours.

In the field of text similarity, corpus-based techniques have solved the most challenging aspect of natural language processing by achieving human-competitive accuracy. However, later paper has shown that even a little difference in text structure or length can easily mislead the prediction. Term frequency inverse document frequency (TF-IDF) is a common approach presented by some that is believed to compensate for the inaccuracy introduced by the document's format and length, but at the expense of

precision. The majority of previous text similarity techniques did not consider the embedding meaning of the words. When working with identical documents other than their wording, the ability of embedding meaning of words becomes useful.

2 LITERATURE REVIEW

PA in medical billing aids the healthcare organization in collecting the correct reimbursement for delivered services, hence lowering denials and subsequent follow-up. According to the results of the Prior Authorization Survey conducted by the American Medical Association (AMA) in December 2021 with the participation of 1,000 practicing physicians, the majority of physicians report an increase in the number of PAs necessary for prescription medications and medical services during the preceding years. In both instances as shown in Fig. 1, the proportion of physicians reporting this rise ranged from 84% to 84% [3].

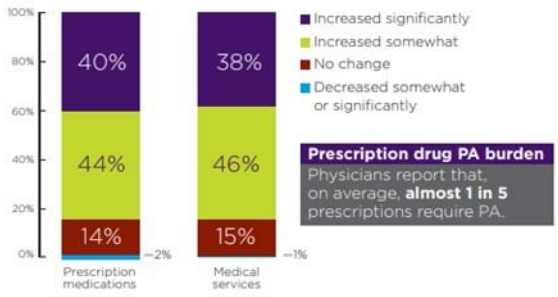

Fig. 1. AMA Prior Authorization Physician Survey [3]

Bag-Of-Words (BOW) [4] and Term Frequency-Inverse Document Frequency (TF-IDF) [5] models have been widely used for text encoding by conventional machine learning algorithms in various text analytics domains, such as the legal sector. Kumar et al. observed that the BOW and TF-IDF model yields better results for recognizing the similarity of legal judgments since just the similarity of the legal phrases, as opposed to all the terms in the dataset, are evaluated [6].

Mandal et al. explored multiple advanced vector representation models for legal document sa uch as Latent Dirichlet Allocation [7] and word embedding along with the TF-IDF model. Word embedding techniques such as Word2vec [8] and Doc2vec [9], that may better capture the semantics of documents, achieved 69% Pearson correlation coefficients in finding similarities among Indian Supreme Court cases, according to the paper [10].

Word embedding representation has been used in the legal field for a number of years. However, the results reported in the reviewed existing literature [10], [11], [12] provide a marginal or non-existent improvement over simple BOW models. This paper

investigates the pre-trained, domain-specific deep learning model utilized for text embedding. After exploring for comparable instances, the newly installed embedding yielded superior results.

3 METHODOLOGY

The method proposes in this work offers an approach for screening PA cases based on local coverage determinations, that are decisions made by a Medicare Administrative Contractor (MAC) whether to approve a case or not. This is an NLP-based model to help insurers by screening the PA cases that give the most similar information and ranking them based on the similarity score in comparison with the Local Coverage Determinations (LCDs) criterions. The Proposed solution consists of the following components as shown in Fig. 2.

1. PA case
2. Requested process id and Healthcare Common Procedure Coding System (HCPCS) code
3. LCDs Criterions
4. Lexicons
5. Text Similarity Check

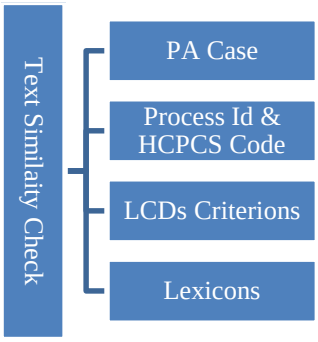

Fig. 2. The Proposed Solution's Components

PA cases are collected through any channel between Insurer and provider, including Electronic Medical Record (EMR). Then, getting prepared three JSON files containing all the necessary information to be utilized as input for the text similarity model.

- The first file contains the rules from the LCD criteria used to approve the procedure in relation to the HCPCS code.
- The second file contains the lexicons associated with each rule; these lexicons aid in the identification of required information from PA cases.
- The third file contains both the requested process id and the HCPCS code needed to verify the eligibility of PA case.

The model retrieves text from the lexicon file and PA case and then executes all of the steps outlined in the following Fig. 3.

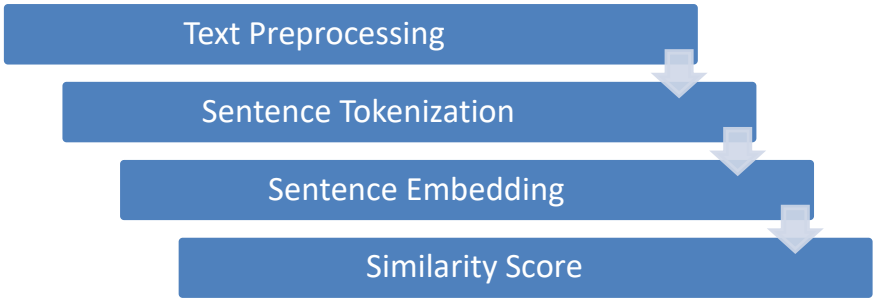

Fig. 3. Text Analysis Workflow

4 SOFTWARE DESIGN

As depicted in Fig. 4, the suggested AI-enabled solution consists primarily of three phases. First, the provider submits a PA request, then the PA automation engine examines the case and give the necessary information from the text to assist Registered Nurses (RNs) in making an approval or denial determination. Now examine the automation engine's underlying framework.

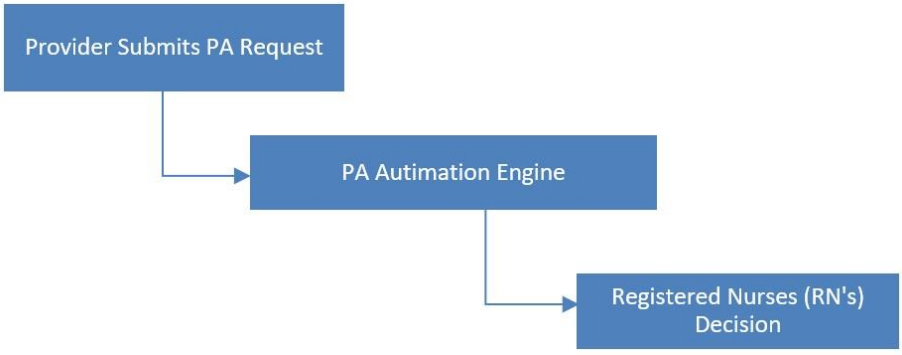

Fig. 4. Proposed AI-Enabled PA Process

There are three stages involved in the operation of an automation engine as shown in Fig. 5. The first stage involves process verification; the second stage identifies rules associated with process id and HCPCS code, as a result, finalizes the lexicons for each rule. and the third stage examines textual similarities. If the process verification phase is successful, the subsequent phases are enabled; if it is failed, a case of rejection is possible. Each step is elaborated on and addressed separately.

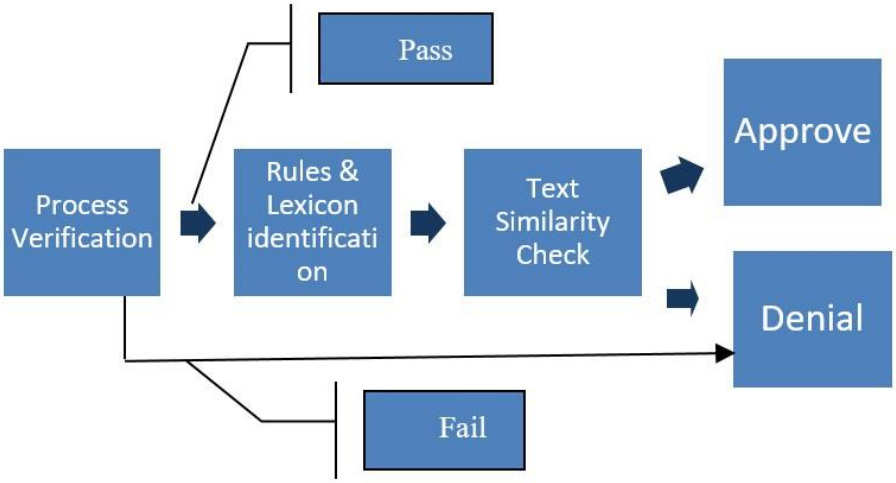

Fig. 5. Automation Engine Process

5 IMPLEMENTATION

In an effort to create an AI-enabled solution. The first and possibly most crucial phase is data preparation because the quality of the solution depends on the quality of the data used. On the payer side, data is collected in accordance with rules. All essential information is kept in JSON format in a configuration folder containing data collected from insurers in accordance with their requirements. The AI-enabled solution for PA requires two components for implementation. The provider provides PA cases in text format with HCPCS codes. Once a PA case containing the requested HCPCS code is received, the automation engine is engaged. According to the design, the automation engine consists of three stages.

The first stage is process verification as shown in Fig. 6 and this procedure begins with the input of the HCPCS code. Therefore, the automation engine employs a rule-based method to cross-check with JSON files containing the process id and corresponding codes. If this eligibility check is successful, the procedure advances to the subsequent level; otherwise, the case is denied.

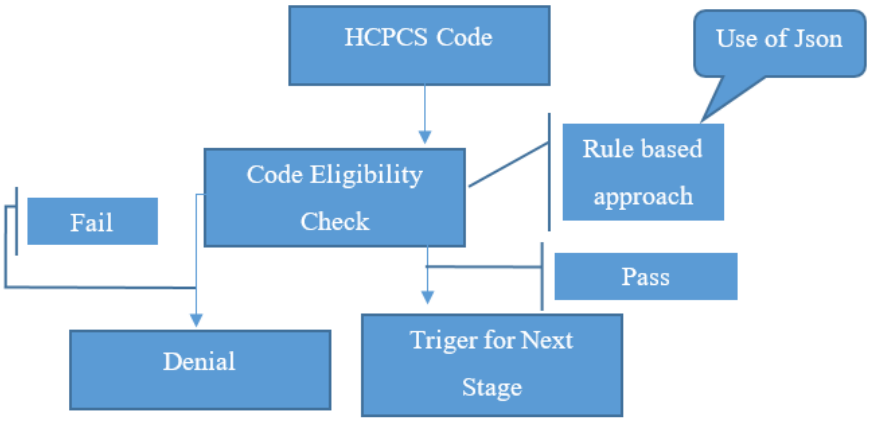

Fig. 6. Process Verification Process

The second stage is rules and lexicon identification. HCPCS code facilitates the identification of rules using a rule-based methodology. Consequently, the lexicon is identified relative to each rule using the same method.

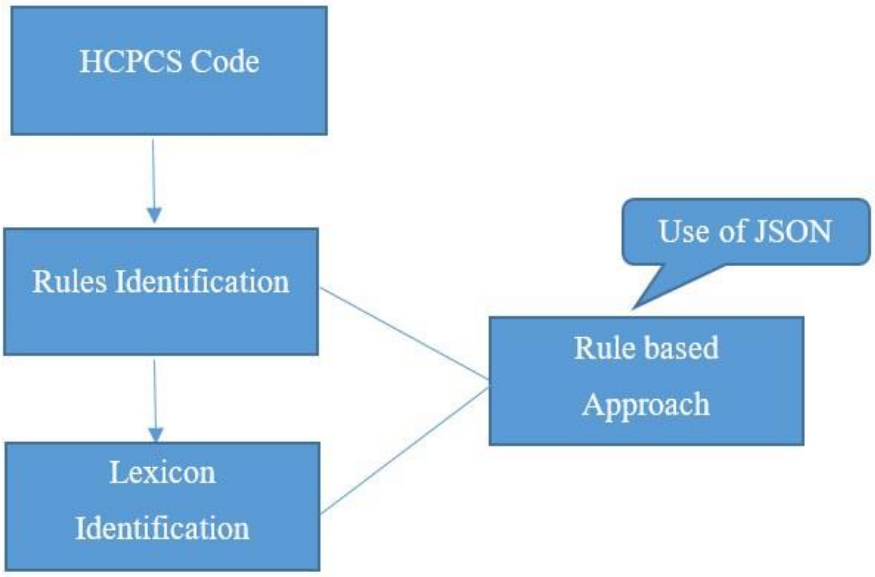

Fig. 7. Rules & Lexicons Identifications Process

The third and last important stage is text similarity. This stage is the central concept of AI utilization in this paper. The text analysis comes into play here.; it aids with text comprehension. Text is provided as input at this stage, after that text preparation occurs. Then, the sentence is tokenized so that the model can convert it into a vector, that is the

sentence embedding process. The identical procedure of sentence embedding was used to the lexicon identified in the previous step. Now, after obtaining two sets of vectors, one from lexicons and the other from text, cosine similarity is used to generate similarity metrics. After receiving ranking evidence, RNs can make a decision based on this evidence.

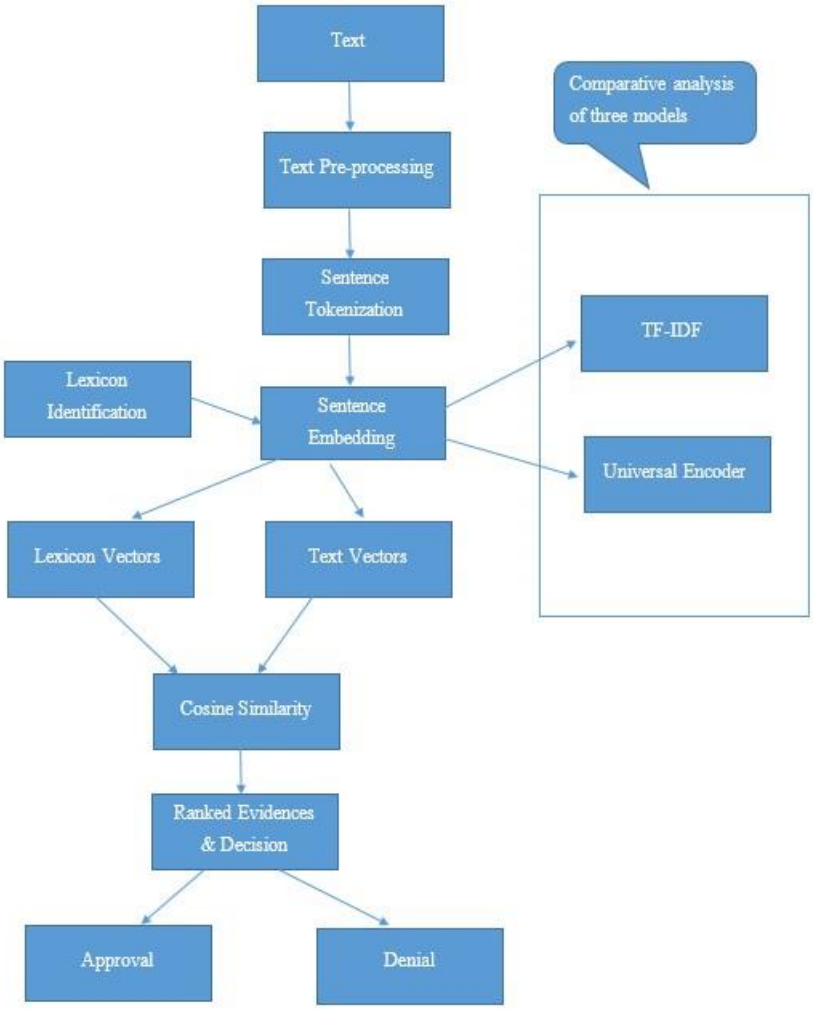

Fig. 8. Text Similarity Process

Sentence embedding is performed to generate a dictionary of pre-processed sentences and their corresponding vectors. Vectors are derived using a model. This model

is founded on NLP principles. In this paper, two NLP strategies were utilized to determine the most effective method for achieving our objective.

5.1 TF-IDF

In this approach, the frequency of words is rescaled according to their overall frequency in all texts, which penalizes common, ubiquitous words like "the" that appear often throughout all texts. TF-IDF measures how crucial a specific term is to the overall meaning of a text. Multiplying two separate metrics yields a document word's TF-IDF. The Term Frequency (TF) of a document's words. There are numerous methods for calculating this frequency, the simplest of that is a simple count of the occurrences of a word in a document. Then, there are further methods for adjusting the frequency. For instance, as Equation (1) describes, by dividing the raw count of occurrences of a word by the document's length or by the raw frequency of the document's most frequent word.

$$TF(i,j)= n(i,j) / \Sigma\, n(i,j) \quad (1)$$

Where,
n(i,j) = number of times nth word occurred in a document
Σ n(i,j) = total number of words in a document.

The inverse document frequency (IDF) of a given word across a collection of documents. This reflects the frequency of a word in the entire document set. The closer a term is to 0, the more frequent it is. This metric can be determined by dividing the total number of documents by the number of documents containing a specific word. This metric can then be calculated using the logarithm.

Therefore, this number approaches 0 if the term is prevalent and appears in several documents. Alternatively, it approaches 1 Multiplying these two numbers yields the TF-IDF score of each word in a document. The higher the score, the more pertinent the word is to the document. In mathematical terms, the TF-IDF score is calculated according to Equation (2).

$$IDF=1+\log(N/dN) \quad (2)$$

Where,
N =Total number of documents in the dataset
dN =total number of documents in that nth word occur

5.2 UNIVERSAL SENTENCE ENCODER

A significant amount of work is expended in machine learning research to convert data into vectors. Word2vec and Glove [13] accomplish this by turning a word into a vector. Therefore, the vector corresponding to "cat" will be closer to "dog" than to "eagle." While embedding a sentence with its words, however, the complete sentence's context

must be captured in that vector. The "Universal Sentence Encoder" comes into play at this point.

The embedding generated by the Universal Sentence Encoder [14] model especially transfer learning to the NLP tasks. It is trained on a number of data sources in order to acquire skills for a vast array of tasks. The sources include Wiki, web media, online question-and-answer pages, and forums. The input is variable-length English text, while the outcome is a 512-dimensional vector.

Typically, sentence embedding was derived by averaging the embedding of all the words in the phrase; however, this method had limitations and was unsuitable for detecting the true semantic meaning of a sentence. The Universal Sentence Encoder makes sentence-level embedding effortless. It is available in two variants, one trained with the Transformer encoder and the other with the Deep Averaging Network (DAN). In terms of computer resource requirements and accuracy, there is a trade-off between the two. While the one with the Transformer encoder is more precise, it requires more computation. The variant with DAN encoding is computationally less expensive and slightly less precise. This paper utilizes the transformer encoder variant.

6 ANALYSIS AND RESULTS

A text similarity-based AI solution must be accurate not only in identifying text but also in identifying the meaning of the word in terms of context, as insurance companies deal with complicated documents as a result of sentence formation in the medical history of patients. As shown by tables 1 and 2, which display the cosine similarity score, a universal sentence encoder thus aids in the capturing of semantic meaning at the sentence level as well as text similarity level. As evident from the validation results, both approaches showed promise in assessing claim approval using text similarity, and reviewers from health insurance companies can approve or decline the claim by confirming that all of the policy's requirements are met in the patient's medical history. This paper shows that, in terms of text similarity, the second approach—using a universal sentence encoder—is more effective than the first—using the TF IDF technique. In this paper, analysis is done by STS Benchmark. This Benchmark offers an empirical assessment of the degree to which similarity ratings obtained by sentence embedding correspond to human judgments. The benchmark necessitates that systems produce similarity scores for an assortment of sentence pairs. The Pearson correlation coefficient is then applied to compare the quality of algorithm similarity scores to human judgments. The statistical method of the Pearson correlation coefficient [15] is commonly used in economics for purposes like trend analysis and classification. Other potential domains of use have been discussed in recent years' literature. Using it, one can determine if strongly two variables are related to one another along a linear axis. And p-value is calculated to check statistically significance. If the correlation coefficient were indeed zero, then the current result would have been seen with a probability equal to the P-value (null hypothesis). A correlation coefficient is considered statistically significant if its associated probability is less than 5%.

6.1 TF-IDF

Pearson correlation coefficient = 0.2340
p-value = 1.015e-19

6.2 UNIVERSAL SENTENCE ENCODER

Pearson correlation coefficient = 0.83
p-value = 0.0

And testing and validation purpose only few rules have been considered that are Rule "A", Rule "B" and Rule "C".

- Rule "A" says one or more mobility-related activities of daily living, are severely hindered because of the beneficiary's mobility limitation.
- Rule "B" says a properly adjusted cane or walker would not help the beneficiary with his/her mobility issues to an acceptable degree.
- Rule "C" says there is insufficient upper extremity function for the beneficiary to propel a properly equipped manual wheelchair indoors.

Results that are from the TF-IDF technique are mentioned in the Table 1.

Table 1. Results from the TF-IDF Technique

Rule Name	Total no. of Matches	Top matching sentence from PA text	Highest Cosine Score
A	5	he limited in his ability to participate in all mobility related activities of daily living in the home setting	0.491431
B	6	he la unable to safely or effectively use cane or walker for the distance needed in the home due to fatigue joint pain and numbness	0.2933
C	9	he is unable to self-propel an optimally configures manual wheelchair due to upper extremity weakness and arthritic hand pain	0.3841

Results that are from the Universal Sentence Encoder Technique are mentioned in the Table 2.

Table 2. Results from the Universal Sentence Encoder Technique

Rule Name	Total no. of Matches	Top matching sentence from PA text	Highest Cosine Score
A	7	he limited in his ability to participate in all mobility related activities of daily living in the home setting	0.72
B	5	he la unable to safely or effectively use cane or walker for the distance needed in the home due to fatigue joint pain and numbness	0.70
C	8	he is unable to self-propel an optimally configures manual wheelchair due to upper extremity weakness and arthritic hand pain	0.71

So, the result from the both technique says Universal Sentence Encoder has a high correlation coefficient, p-value, and cosine score compared to TF_IDF.

The similarity is visualized using a heat map. The graph is a 31x52 matrix and the color of each element [i, j] is determined by the dot product of the embedding for sentence i and j. The similarity between sentences is displayed via a heat map. This result demonstrates that because the embedding process is carried out at the sentence level, the outcomes from the second approach using the Universal Sentence Encoder have a high color intensity. The two documents being compared are one a medical history and the other a set of policy guidelines. Figure 6 shows the heat map for TF-IDF and Figure 7 shows the heat map for Universal Sentence Encoder.

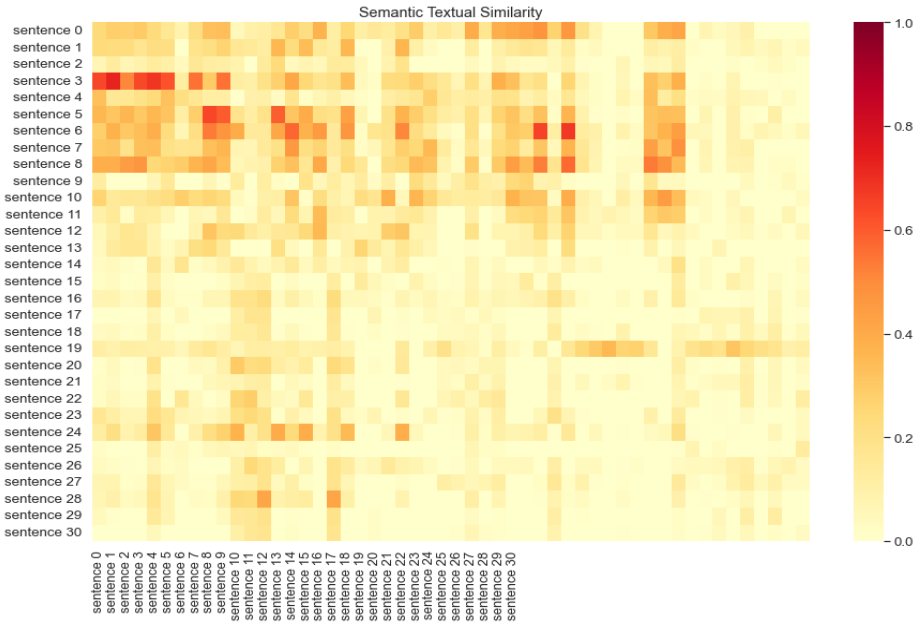

Fig. 9. Similarity Visualization for TF-IDF

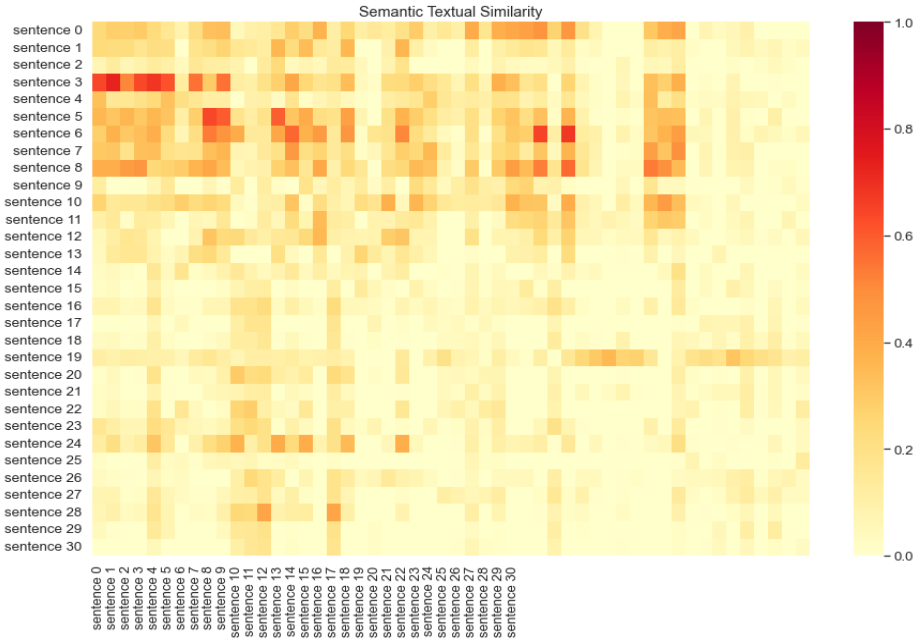

Fig. 10. Similarity visualization for Universal Sentence Encoder

7 CONCLUSION

The facts presented in the previous section make it plainly clear that Universal Sentence Encoder's tactics are superior to those of TF IDF. Using the Pearson correlation coefficient, the assessment standard allows for differentiation between the various methodologies. The similarity visualization generates outputs with varying degrees of color intensity that are quite similar. This conclusion is based on the information gathered for the purpose of this paper. On the other hand, the evaluation benchmark makes it obvious that Universal Sentence Encoder approaches can still tackle the challenge even if the difficulty of the text increases. Even though this paper is conducted on PA cases, that are part of the healthcare industry, complexity management is always a concern. This methodology aids in reaching the objectives of the paper. Access to necessary care for patients is frequently delayed as a result of prior authorizations, that may drive patients to abandon their treatment due to the waiting period or other complications related with prior authorization. This work provides a viable method for resolving the issue, as it proposes a method for streamlining AI that minimizes treatment delays and disruptions by reducing the requirement for prior approval. This solution, that is an integral part of the End-to-End Prior Authorization process, eliminates human work that is time-consuming and prone to error. Therefore, the pre authorization team can maximize the health system's capacity to provide faster and better care. Patients, healthcare providers, and insurers, as well as any other parties engaged in the process, can all benefit from an efficient utilization management program. These are the adverbial complements for each:

- Patients gain from decreased treatment costs, more treatment efficacy, and fewer refused claims.
- Fewer denied claims, reduced costs, more effective treatments, improved data, and more efficient resource utilization are all beneficial to the health care industry.

References

1. T. M. Wickizer and D. Lessler, “Utilization management: Issues, effects, and future prospects,” Annu. Rev. Public Health, vol. 23, pp. 233–254, 2002, doi: 10.1146/ANNUREV.PUBLHEALTH.23.100901.140529..
2. “AI ushers in next-gen prior authorization in healthcare | McKinsey | McKinsey.” https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/ai-ushers-in-next-gen-prior-authorization-in-healthcare (accessed Aug. 10, 2022).
3. A. Medical Association, “Prior Authorization Physician Survey Update | AMA,” 2022, Accessed: Aug. 10, 2022. [Online]. Available: https://www.ama-assn.org/system/files/prior-.
4. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” Decis. Support Syst., vol. 50, no. 3, pp. 559–569, 2011, doi: 10.1016/J.DSS.2010.08.006.

5. J. Lam, Y. Chen, F. Zulkernine, and S. Dahan, “Detection of Similar Legal Cases on Personal Injury,” IEEE Int. Conf. Data Min. Work. ICDMW, vol. 2021-December, pp. 639–646, 2021, doi: 10.1109/ICDMW53433.2021.00084.
6. M. Kumar, R. Ghani, and Z. S. Mei, “Data mining to predict and prevent errors in health insurance claims processing,” Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 65–73, 2010, doi: 10.1145/1835804.1835816.
7. D. Blei, A. Ng, M. J.-J. of machine L. research, and undefined 2003, “Latent dirichlet allocation,” jmlr.org, vol. 3, pp. 993–1022, 2003, Accessed: Aug. 10, 2022. [Online]. Available: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=https://githubhelp.com.
8. T. Mikolov, K. Chen, G. Corrado, J. D. preprint arXiv:1301.3781, and undefined 2013, “Efficient estimation of word representations in vector space,” arxiv.org, Accessed: Aug. 10, 2022. [Online]. Available: https://arxiv.org/abs/1301.3781.
9. Q. Le, T. M.-I. conference on machine, and undefined 2014, “Distributed representations of sentences and documents,” proceedings.mlr.press, Accessed: Aug. 10, 2022. [Online]. Available: http://proceedings.mlr.press/v32/le14.html?ref=https://githubhelp.com.
10. A. Mandal, R. Chaki, S. Saha, K. Ghosh, A. Pal, and S. Ghosh, “Measuring similarity among legal court case documents,” ACM Int. Conf. Proceeding Ser., pp. 1–9, Nov. 2017, doi: 10.1145/3140107.3140119.
11. C. Xia, T. He, W. Li, Z. Qin, and Z. Zou, “Similarity Analysis of Law Documents Based on Word2vec,” Proc. - Companion 19th IEEE Int. Conf. Softw. Qual. Reliab. Secur. QRS-C 2019, pp. 354–357, Jul. 2019, doi: 10.1109/QRS-C.2019.00072.
12. D. Thenmozhi, K. Kannan, C. A.-F. (Working Notes), and undefined 2017, “A text similarity approach for precedence retrieval from legal documents.,” ceur-ws.org, Accessed: Aug. 10, 2022. [Online]. Available: http://ceur-ws.org/Vol-2036/T3-9.pdf.
13. J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation,” Accessed: Aug. 15, 2022. [Online]. Available: http://nlp.
14. D. Cer et al., “Universal Sentence Encoder,” AAAI, pp. 16026–16028, Mar. 2018, doi: 10.48550/arxiv.1803.11175.
15. X. Zhi, S. Yuexin, M. Jin, Z. Lujie, and D. Zijian, Research on the Pearson correlation coefficient evaluation method of analog signal in the process of unit peak load regulation.