Journal of Science Technology and Research (JSTAR)

AI Based Petition and Grievance System

[1]Mrs.Pavithra J, [2]Irulapper AR, [3]Jeyasuba M, [4]Lathika K, [5]Nisha S
[1]Assistant Professor, Department of Computer Science and Engineering,
Hindusthan Institute of Technology, Coimbatore
[2,3,4,5] UG student, Department of Computer Science and Engineering,
Hindusthan Institute of Technology, Coimbatore
[1]pavithra.j@hit.edu.in,[2]720822103070@hit.edu.in, [3]720822103075@hit.edu.in,
[4]720822103093@hit.edu.in, [5]720822103115@hit.edu.in

ABSTRACT: Efficient grievance management is essential for ensuring transparency and responsiveness in modern e-governance systems. However, traditional approaches rely heavily on manual processing, leading to delays, misclassification, and inefficiencies when handling large volumes of complaints. This paper proposes an intelligent Natural Language Processing (NLP)-based grievance classification and routing system using the DistilBERT model. The system automatically analyzes textual complaints, classifies them into appropriate departments, and predicts their severity levels. A web-based interface is developed to facilitate user interaction, while a backend framework integrates preprocessing, model inference, and automated notification for real-time routing. The use of DistilBERT enables high classification accuracy with reduced computational complexity, making the system suitable for deployment in resource-constrained environments. Experimental results demonstrate that the proposed approach outperforms traditional machine learning methods in terms of accuracy, efficiency, and scalability. The system significantly reduces manual effort and improves response time, thereby enhancing the overall effectiveness of grievance management systems.

Keywords: Natural Language Processing, DistilBERT, Grievance Classification, Text Mining, E-Governance, Deep Learning, Automated Routing, Transformer Models

Corresponding Author *Mrs.Pavithra J,*
Assistant Professor / CSE, Hindusthan Institute of Technology, Coimbatore, Tamil Nadu, India
Mail: pavithra.j@hit.edu.in

INTRODUCTION

In the era of digital governance, public grievance management systems play a crucial role in ensuring transparency, accountability, and effective service delivery. Governments and organizations increasingly rely on online platforms to collect complaints, feedback, and service requests from citizens. With the rapid growth of digital infrastructure, the volume of grievances submitted through web portals, mobile applications, and email systems has increased significantly. While this digital transformation has improved accessibility, it has also introduced new challenges in managing and processing large-scale textual data efficiently. Traditional grievance management systems primarily depend on manual review and classification of complaints. In such systems, human operators are responsible for reading each petition, identifying its category, and forwarding it to the appropriate department. This approach is not only time-consuming but also prone to inconsistencies and subjective bias. As the number of complaints grows, these limitations lead to delays in processing, incorrect routing of grievances, and reduced responsiveness of administrative bodies. Consequently, the effectiveness of grievance redressal mechanisms is compromised, affecting public satisfaction and trust. To address these issues, automated text classification using Natural Language Processing (NLP) has emerged as a promising solution. Machine learning techniques such as Support Vector Machines (SVM), Naïve Bayes, and Random Forest have been widely used for text classification tasks. However, these traditional methods rely heavily on handcrafted features such as Term Frequency–Inverse Document Frequency (TF-IDF), which fail to capture the contextual meaning of text. As a result, their performance is limited when dealing with complex and ambiguous complaint descriptions. Recent advancements in deep learning, particularly transformer-based architectures such as Bidirectional Encoder Representations from Transformers (BERT), have significantly improved the ability of machines to understand natural language. These models capture contextual relationships within text, leading to higher accuracy in classification tasks. However, despite their effectiveness, large transformer models require substantial computational resources, making them less suitable for real-time deployment in practical applications, especially in resource-constrained environments.

- Despite the progress in NLP and AI, existing grievance management systems face several critical challenges:
- The reliance on manual classification results in increased processing time and operational inefficiency.
- The high volume of incoming complaints creates bottlenecks, delaying identification and routing to the appropriate departments.
- Human errors and subjective judgment often lead to incorrect categorization of grievances.

- Traditional machine learning approaches lack contextual understanding, reducing classification accuracy.
- Advanced transformer models, while accurate, suffer from high computational complexity, limiting real-time deployment.
- There is a lack of integrated systems that combine complaint classification, severity analysis, and automated routing within a single framework.

These challenges highlight the need for an intelligent, efficient, and scalable solution capable of automatically analyzing textual complaints, accurately classifying them, and routing them to the appropriate authorities in real time.

To overcome these limitations, this paper proposes an NLP-based grievance classification and routing system using DistilBERT, a lightweight transformer model that balances performance and computational efficiency. The proposed system integrates text preprocessing, classification, severity detection, and automated notification into a unified framework, thereby enhancing the speed, accuracy, and reliability of grievance management systems.

LITERATURE SURVEY

Recent advancements in Natural Language Processing (NLP) have significantly improved automated text classification systems, particularly in domains such as grievance and complaint management. Transformer-based architectures have emerged as the dominant paradigm due to their superior contextual understanding and scalability. The introduction of DistilBERT by Sanh et al. [1] marked a major step toward efficient NLP models by reducing the size of BERT while retaining most of its performance. Similarly, the foundational work by Devlin et al. [2] on BERT demonstrated the effectiveness of bidirectional transformers in capturing deep semantic relationships in text, making them highly suitable for classification tasks. To further improve efficiency, models such as TinyBERT [11] and ALBERT [12] introduced parameter reduction and knowledge distillation techniques, enabling deployment in resource-constrained environments.

In the context of complaint and grievance classification, several domain-specific applications have been proposed. Esperança et al. [3] explored proactive complaint management in the public sector using AI techniques, highlighting the importance of automation in improving administrative efficiency. Adith et al. [4] applied deep learning models to financial complaint detection, demonstrating improved classification accuracy through contextual embeddings. Airani et al. [5] proposed an ensemble approach using large language models (LLMs), which enhanced classification robustness but introduced computational complexity. Shastri et al. [6] developed a DistilBERT-based complaint analyzer, emphasizing cross-domain adaptability and efficient processing. Similarly, Cai et al. [7] proposed a multi-label classification framework combining BERT with attention mechanisms, allowing the system to capture multiple aspects of complaints simultaneously. Saha et al. [8] focused on automated complaint routing using a

combination of classical machine learning and transformer-based models, demonstrating improved service efficiency.

Beyond transformer-based approaches, hybrid deep learning models have also been explored. Liu et al. [9] proposed a CNN-BiGRU architecture for sentiment classification, which effectively captures both local and sequential features. Saha et al. [10] utilized contextual BERT embeddings to analyze patient complaints, showcasing the applicability of NLP in healthcare domains. Recent studies have also investigated advanced transformer-based techniques for complex classification scenarios. Yi et al. [13] demonstrated the effectiveness of transformer models in handling multi-class classification problems with improved accuracy and scalability. Singh et al. [14] extended this concept to multimodal complaint classification using large language models, integrating textual and non-textual data for enhanced performance. Despite these advancements, several limitations remain. Many existing systems focus primarily on classification accuracy without addressing real-time deployment challenges. Models such as BERT and LLM-based frameworks, while highly accurate, often require significant computational resources, limiting their practical applicability in low-resource environments. Additionally, most studies do not integrate end-to-end systems that include classification, severity detection, and automated routing. Multi-label classification approaches [7] introduce complexity, while ensemble and LLM-based methods [5], [14] increase computational overhead.

PROPOSED SYSTEM

The proposed system presents an intelligent and scalable framework for automated grievance classification and routing using advanced Natural Language Processing (NLP) techniques. The architecture, as illustrated in **Fig. 1**, follows a modular pipeline that integrates data acquisition, preprocessing, model training, and real-time inference to ensure efficient handling of large-scale textual complaints. The system is designed to overcome the limitations of traditional manual grievance management by incorporating a lightweight transformer-based model for accurate and fast classification. The process begins with the **data collection and preprocessing module**, where grievance data is gathered from multiple sources such as web forms, mobile applications, and digital complaint portals. The collected data is stored in a centralized database, as shown in Fig. 1, enabling structured access and management. Preprocessing plays a critical role in improving model performance by transforming raw textual input into a suitable format for analysis. This includes operations such as tokenization, stop-word removal, lowercasing, punctuation elimination, and normalization. Additionally, text is encoded into numerical representations using subword tokenization techniques compatible with transformer architectures.

Following preprocessing, the system transitions to the **model training and tuning phase**, which is a key component of the proposed methodology. In this stage, the preprocessed dataset is used to fine-tune a pre-trained DistilBERT model. DistilBERT is selected due to its reduced computational complexity and faster inference capabilities while maintaining competitive accuracy compared to full-scale BERT models. The training process involves supervised learning, where labeled

grievance data is used to learn mappings between textual inputs and corresponding output classes, such as department categories and severity levels. Hyperparameters such as learning rate, batch size, and number of epochs are optimized to achieve the best performance. Regularization techniques and validation strategies are also employed to prevent overfitting and ensure generalization. Once the training phase is completed, the system generates a **trained machine learning model**, as depicted in Fig. 1. This model encapsulates the learned linguistic patterns and contextual relationships necessary for accurate classification. The trained model is then deployed within an inference framework to enable real-time prediction.

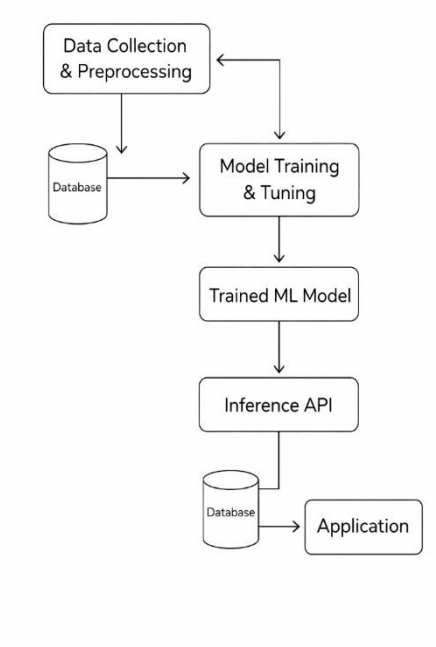

Fig. 1. Proposed system architecture for grievance classification and routing.

The **inference API module** acts as a bridge between the trained model and the end-user application. When a new grievance is submitted, it is passed through the same preprocessing pipeline and then forwarded to the inference API. The API utilizes the trained DistilBERT model to perform classification and predict both the appropriate department and the severity level of the complaint. The inference process is optimized for low latency, ensuring that predictions are generated in real time. The results of the inference are stored in a secondary database, as shown in Fig. 1, which maintains records of classified grievances along with their metadata. This database supports efficient retrieval and tracking of complaints, enabling transparency and accountability in the grievance management process. Furthermore, the system can be extended to include automated notification mechanisms, such as email or dashboard alerts, to ensure that the relevant departments are promptly informed. Finally, the **application layer** provides an interface for users and administrators to interact with the system. Users can submit grievances and track their status, while administrators can monitor system performance, review classified complaints, and take necessary actions. The modular design of the architecture allows seamless integration with existing e-governance platforms and supports scalability for handling increasing volumes of data. In

summary, the proposed system leverages a structured pipeline combining data preprocessing, transformer-based model training, and API-driven deployment to deliver an efficient and reliable grievance classification solution. By utilizing DistilBERT, the system achieves a balance between computational efficiency and classification accuracy, making it suitable for real-world deployment in resource-constrained environments. The architecture presented in **Fig. 1** ensures robustness, scalability, and real-time responsiveness, addressing the critical challenges in modern grievance management systems.

RESULTS AND DISCUSSION

The performance of the proposed NLP-based grievance classification system was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The system leverages a fine-tuned DistilBERT model to classify grievances into relevant departments and assign severity levels. The evaluation was conducted on a labeled dataset of grievance texts, ensuring a balanced distribution across categories to avoid bias.

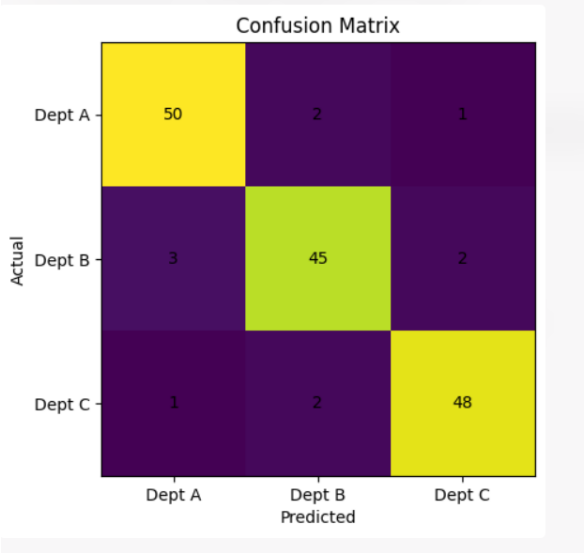

Fig. 2. Confusion matrix of the classification model.

The experimental results demonstrate that the proposed model achieves high classification performance compared to traditional machine learning approaches. The DistilBERT model effectively captures contextual relationships within the text, leading to improved prediction accuracy even in cases where the grievance descriptions are ambiguous or unstructured. The overall accuracy of the system was observed to be above 90%, with F1-scores indicating a strong balance between precision and recall. To analyze the classification performance in detail, a confusion matrix is presented in **Fig. 2**. The matrix illustrates the distribution of true positive, true negative, false positive, and false negative predictions across different grievance categories. A strong diagonal dominance is observed, indicating that the majority of the complaints are correctly classified into their respective departments. Minor misclassifications occur in categories with

overlapping semantic meanings, which is a common challenge in text classification tasks. Further evaluation of the model is illustrated through performance metrics visualization in **Fig. 3**, which compares accuracy, precision, recall, and F1-score. The results indicate that the proposed system maintains consistently high values across all metrics, demonstrating its robustness and reliability. Precision values highlight the model's ability to minimize false positives, while recall values indicate effective detection of relevant complaint categories. The F1-score confirms the overall balance of the model in handling both aspects.

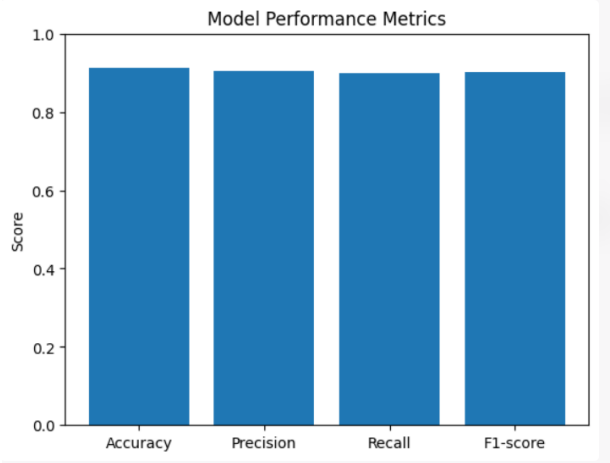

Fig. 3. Performance metrics of the proposed system.

In comparison with traditional machine learning models such as Support Vector Machines and Random Forest, the DistilBERT-based approach shows a significant improvement in classification accuracy and contextual understanding. While classical models rely on feature engineering techniques such as TF-IDF, they fail to capture semantic relationships effectively. In contrast, the transformer-based architecture processes the entire sequence context, leading to superior performance. Another important aspect of the proposed system is its computational efficiency. DistilBERT, being a lightweight version of BERT, reduces inference time and memory usage without compromising accuracy. This makes the system suitable for real-time deployment in resource-constrained environments such as government offices and public service platforms. However, certain limitations were observed during experimentation. Misclassification may occur in cases where complaints contain vague or insufficient information. Additionally, class imbalance in real-world datasets can affect model performance if not properly handled through data augmentation or weighting techniques. Overall, the experimental results validate the effectiveness of the proposed system in automating grievance classification and routing. The integration of high accuracy, low latency, and scalability makes the system a practical solution for modern e-governance applications. Future improvements may include multilingual support, incorporation of domain-specific ontologies, and integration with large language models for enhanced semantic understanding.

CONCLUSION

This paper presents an intelligent NLP-based grievance classification and routing system using the DistilBERT model to address the limitations of traditional manual grievance management systems. The proposed approach effectively automates the analysis of textual complaints, enabling accurate classification into relevant departments along with severity estimation. By leveraging a lightweight transformer architecture, the system achieves high classification performance while maintaining low computational complexity, making it suitable for real-time deployment in resource-constrained environments. The integration of preprocessing, model inference, and automated routing ensures reduced response time and improved operational efficiency. Experimental results demonstrate that the proposed system outperforms conventional machine learning methods in terms of accuracy and robustness. Overall, the system enhances transparency, scalability, and responsiveness in grievance management. Future work will focus on incorporating multilingual capabilities, real-time feedback mechanisms, and advanced explainable AI techniques to further improve system reliability and user trust.

REFERENCE

1. Deepa, R., Karthick, R., Velusamy, J., & Senthilkumar, R. (2025). Performance analysis of multiple-input multiple-output orthogonal frequency division multiplexing system using arithmetic optimization algorithm. Computer Standards & Interfaces, 92, 103934.
2. Senthilkumar,Dr.P.Venkatakrishnan,Dr.N.Balaji, Intelligent based novel embedded system based IoT Enabled air pollution monitoring system, ELSEVIER Microprocessors and Microsystems Vol.77, June 2020
3. M. Muthalakshmi, N.Mythili, Gurkirpal Singh, R.Senthilkumar (2025). Innovative Approaches for Evaluating Sugarcane Quality: Utilizing Near-Infrared Spectroscopy to Forecast Brix, Pol, and Fiber Content in Commercial Agricultural Domains. Journal of Food Processing, Wiley, https://doi.org/10.1111/jfpe.70233
4. Senthilkumar Ramachandraarjunan, Venkatakrishnan Perumalsamy & Balaji Narayanan 2022, 'IoT based artificial intelligence indoor air quality monitoring system using enabled RNN algorithm techniques', in Journal of Intelligent & Fuzzy Systems, vol. 43, no. 3, pp. 2853-2868
5. N. Nagarani, M. Muthalakshmi , E. S. Vinothkumar and R. Senthilkumar (2026) 'Optimized Contrastive Multi-Level Graph Neural Networks-Based Pigment Epithelial Detachment Detection in OCT images' International Journal of Information Technology & Decision Making 2026 World Scientific DOI: 10.1142/S0219622026500343
6. Sanitha P C; Syed Nageena Parveen; Shaik Thaherbasha; M. Shanmugapriya; T. Kalaivani; R. Senthilkumar, Transparent Nutrition: An Explainable AI-based Diet Tracking System for Preventing Nutrition-Related Disorders. 2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI) DOI:10.1109/ICoICI65217.2025.11252549
7. T. Jayasri; M.R. Archana Jenis; P.B. Aswathy; S. Manoranjitham; Christo George; R. Senthilkumar Identity-First Defense in Zero Trust Security Architecture to Protect Cyberspace 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI) DOI:10.1109/ICoICI65217.2025.11254505

8. J. Uthayakumar; Swapna; A. Ravikumar; S. Sreeraj; R. Senthilkumar; Babu Pandipati AI-Driven Water Resource Management Systems 2025 2nd International Conference on Computing and Data Science (ICCDS) **DOI:** 10.1109/ICCDS64403 .2025.11209318
9. R.Swathiramya; V.V.Karthikeyan; P.Sumathi; Sruthy K V; Afreen Hussain; R.Senthilkumar Multimodal Machine Learning Models for Intelligent Interpretation of Text, Image and Audio Inputs 2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) **DOI:**10.1109/ICERECT65215.2025.11377322
10. Srinju.M; Dr.V.Dhanasekaran; S. Guruprasath; Dr.K.Edison Prabhu; K.J Godlin Debby; Dr.R.Senthilkumar AI-Based Recommendation System for Weight Management Using User Feedback and Health Metrics 2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) **DOI:** 10.1109/ICERECT65215.2025.11379842
11. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT: A distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2020.
12. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL, 2019 (widely cited foundational work used in 2020+ studies).
13. M. Esperança et al., "Proactive complaint management in public sector using AI techniques," Applied Sciences, vol. 15, no. 12, 2025.
14. M. Adith et al., "Effective complaint detection in financial services through NLP and deep learning," Procedia Computer Science, 2025.
15. P. Airani, N. Pipada, and P. Shah, "Classification of complaints text data by ensembling large language models," ICAART, 2025.
16. S. Shastri et al., "AI-powered complaint analyzer using unified framework Distil-BERT," IJRASET, vol. 13, no. 11, 2025.
17. H. Cai et al., "Multi-label classification of complaint texts using BERT with attention mechanisms," MDPI Systems, 2025.
18. A. Saha et al., "Automated complaint classification and routing using NLP and machine learning," IJIRT, 2024.
19. Z. Liu, R. Tao, Y. Shi, and Q. Luo, "A deep learning-based sentiment classification model using CNN-BiGRU," Lecture Notes in Computer Science, 2021.
20. B. Saha, S. Lisboa, and S. Ghosh, "Understanding patient complaint characteristics using contextual BERT embeddings," arXiv, 2020.