Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

LLM-Driven Semantic Understanding and Automatic Categorization of Public Issue Reports in Smart Cities

Hitesh Bhor, Bhushan Kumavat, Tilak Jain, Vedant Shirsath,
Prof. P. G. Fegade, Prof. P. P. Patil, Prof. A. L. Rane
Master of Computer Applications
K.K. Wagh Institute of Engineering Education and Research
Nashik, India

***Abstract*—The rapid digitalization of public services has led to a sharp increase in citizen-submitted issue reports, which are typically unstructured, multilingual, and multimodal in nature. Manual categorization of such reports is neither scalable nor consistent, resulting in delayed grievance resolution and poor resource utilization. This paper presents Jagruk, an LLM-driven pipeline for the semantic understanding, automatic classification, and priority-based routing of civic complaint reports in smart city environments. The proposed system accepts unstructured text along with optional image and geolocation inputs, and processes them through a multimodal natural language processing workflow using instruction-tuned Large Language Models (LLMs). Each complaint is classified into a predefined municipal category, assigned an urgency priority, and routed to the appropriate department, all within a single inference step. The system is evaluated against three baseline models—TF-IDF with SVM, Multinomial Naive Bayes, and fine-tuned BERT—on a labeled civic complaint dataset. The proposed LLM-based system achieves an overall accuracy of 94.2%, a macro F1-score of 92.9%, and a Mean Absolute Error (MAE) of 0.54 for priority prediction, outperforming all baselines. Deployment considerations including inference latency, data privacy, and fairness are also discussed, demonstrating the suitability of the system for real-world e-governance applications.**

***Index Terms*—Large Language Models, semantic analysis, smart city, e-governance, complaint classification, natural language processing, grievance management, multimodal processing**

I. Introduction

The rapid digitalization of public services has resulted in a substantial increase in the volume of citizen-submitted issue reports that require timely processing and categorization. Manual methods of complaint management are not scalable and are prone to delays, inconsistencies, and inefficient resource allocation [1].

In this paper, we propose Jagruk, an LLM-driven framework for the semantic understanding and automatic categorization of public issue reports in smart cities. The system harnesses Large Language Models to parse unstructured and multimodal inputs, and accurately classifies, prioritizes, and routes complaints to the relevant municipal departments [2]–[4].

Unlike conventional keyword-based or rule-based systems, the proposed framework captures contextual meaning, thereby accommodating informal language, code-mixed text, and multilingual citizen reports [5]. This significantly enhances the efficiency, transparency, and responsiveness of e-governance platforms.

A. Problem Statement

Citizen complaints predominantly consist of unstructured, noisy text containing informal vocabulary and code-mixed language, with optional multimedia inputs such as images and location data. Manual processing of such heterogeneous inputs is error-prone, inconsistent, and fails to scale with growing complaint volumes. The primary challenge addressed in this work is the design of an automated system capable of accurately understanding, classifying, and prioritizing civic complaints while preserving fairness and protecting user privacy.

B. Key Contributions

The principal contributions of this work are as follows:

- A multimodal LLM-based pipeline that jointly processes text descriptions, images, and geolocation data for complaint understanding.
- A single-inference classification and routing mechanism that assigns complaint category, urgency priority, and responsible department simultaneously.
- A priority detection mechanism driven by sentiment and urgency cues, coupled with a community voting system to surface widely reported issues.
- A real-time, end-to-end civic complaint management platform (Jagruk) deployed with a full-stack web architecture and integrated with municipal workflow services.
- Empirical evaluation demonstrating significant performance improvements over TF-IDF+SVM, Naive Bayes, and fine-tuned BERT baselines.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

C. Importance in Smart Cities and E-Governance

Automated complaint classification reduces response times, increases operational efficiency, and supports data-driven decision-making in municipal governance. By enabling scalable management of high-volume complaint streams, the system promotes transparency between government and citizens and improves the quality of public service delivery.

D. Challenges in Manual Categorization

Existing manual systems are slow, inconsistent, and incapable of handling the linguistic diversity of citizen-generated content. They do not scale with growing complaint databases, and their inability to interpret contextual or multimodal data results in frequent misrouting and diminished service efficacy.

II. Literature Review

A. Existing Complaint Management Systems

Classical complaint management systems in e-governance rely on manual sorting or rule-based routing. These systems enforce rigid workflows that lack flexibility and become ineffective as complaint volume and diversity increase [6], [7]. Earlier systems were unable to automatically respond to unstructured complaints, leading to persistent delays and unresolved grievances. More recent attempts at automation still struggle with contextual understanding, particularly in multilingual settings [5]. Unlike these systems, the proposed approach requires no predefined rules and handles informal, code-mixed text through LLM-based semantic inference.

B. NLP-Based Approaches

Natural language processing has been widely applied to complaint classification. Early approaches employed TF-IDF with N-gram features and machine learning classifiers such as Naive Bayes and SVM. These methods require extensive manual feature engineering and perform poorly on informal, noisy, or semantically ambiguous text. The introduction of deep learning models, particularly LSTMs with attention mechanisms, improved classification accuracy. Transformer-based models such as BERT have further advanced semantic classification capabilities [1], [3], establishing new performance benchmarks across multiple complaint classification tasks. However, BERT-based systems require substantial labeled data for fine-tuning and do not generalize well to out-of-distribution complaint categories without retraining [9].

C. Machine Learning versus Large Language Models

Standard machine learning models are computationally efficient but exhibit limited semantic understanding and contextual adaptability. Large Language Models, by contrast, support zero-shot and few-shot classification [4], enabling robust handling of diverse and evolving complaint categories with minimal supervision. Recent work has demonstrated that instruction-tuned LLMs have been shown to outperform in several short-text classification scenarios, particularly when training data is scarce or categories are loosely defined [11]. However, LLMs introduce challenges related to computational cost and inference latency. Recent research indicates that hybrid approaches combining LLM reasoning with lightweight traditional models can achieve a favorable balance between accuracy and efficiency [8], [9], [13]–[15]. Furthermore, multimodal extensions of transformer architectures have shown that incorporating visual context alongside text consistently improves classification performance in ambiguous scenarios [12].

D. Comparative Summary

Table I summarizes the key characteristics of representative approaches in the literature. The proposed system advances beyond prior work by combining multimodal inputs, zero-shot LLM classification, and integrated priority estimation within a single deployable platform.

TABLE I
Comparison of Existing Complaint Classification Approaches

Approach	Multimodal	Multilingual	Priority	Accuracy
TF-IDF + SVM [7]	No	No	No	∼76%
BERT Fine-tuned [1]	No	Limited	No	∼86%
RailNeural [6]	No	No	No	∼83%
Zero-shot LLM [4]	No	Yes	No	∼88%
Proposed (Jagruk)	**Yes**	**Yes**	**Yes**	**94.2%**

III. Proposed System

Jagruk is a web-based civic complaint management platform designed for smart city environments, where the high volume, diversity, and unstructured nature of citizen reports renders manual sorting impractical. Unlike traditional rule-based approaches, Jagruk employs an instruction-tuned Large Language Model for semantic understanding and autonomous decision-making.

Every submitted complaint is processed by the LLM, which generates structured metadata encompassing the issue category, estimated urgency, responsible department, and a concise rationale. This metadata drives automated routing and prioritization. The platform is built on a full-stack architecture comprising React for the frontend, Node.js/Express for backend orchestration, MongoDB for persistent storage, and several external microservices for media handling, authentication, geolocation, and LLM access.

A. System Architecture

Jagruk employs a modular, multi-tier architecture organized into four primary layers, as illustrated in Fig. 1.

Presentation Layer: A React single-page application provides role-based interfaces for citizens and administrators. Citizens submit complaints with optional image and location data, while administrators access dashboards for complaint tracking, queue management, and analytics.

Service Layer: Built with Node.js and Express, this layer exposes RESTful APIs connecting the frontend to backend services. It handles input validation, orchestration logic, and integration with external services.

Intelligence Layer: An instruction-tuned Large Language Model, accessed via the OpenRouter API, classifies structured

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

prompts into JSON-formatted outputs for automated categorization and prioritization.

Persistence Layer: MongoDB stores enriched complaint records together with full audit trails, enabling traceability and compliance monitoring.

External microservices include Appwrite for authentication and role management, Cloudinary for image storage and delivery, OpenCage for reverse geocoding, and OpenRouter for LLM access.

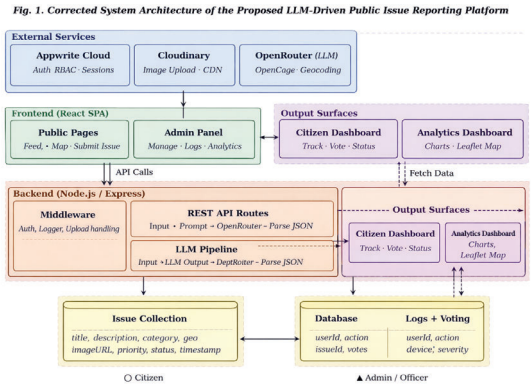

Fig. 1. System architecture of the Jagruk LLM-driven civic issue platform.

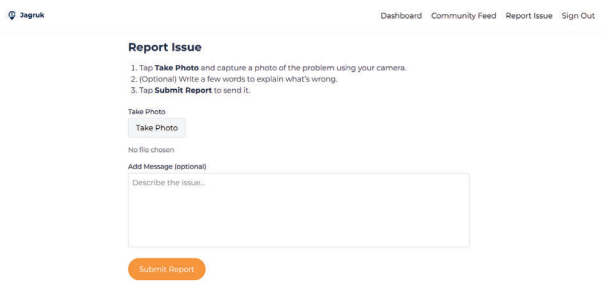

Fig. 2. Citizen complaint submission interface in Jagruk.

B. Input Modalities

The system accepts three categories of input: a natural language complaint description (mandatory), an optional image, and optional geolocation data. The text field imposes no length or language restriction, accommodating reports ranging from brief phrases to detailed descriptions, including multilingual and code-mixed content.

When an image is present, it is uploaded to Cloudinary and its CDN URL is incorporated into the processing pipeline as additional visual context. When location access is granted by the user, the browser's Geolocation API provides latitude and longitude coordinates, which are converted to a human-readable address using the OpenCage reverse geocoding service. All three inputs are combined into a unified context payload by the backend before being forwarded for LLM processing.

C. LLM-Based Processing

Upon receiving a validated input payload, the backend constructs a structured prompt comprising two components.

The *system prompt* specifies the classification task, defines the permitted complaint categories and urgency levels, and mandates a strict JSON output schema with the fields: `category`, `priority`, `rationale`, and `department`.

The *user prompt* contains the dynamically populated complaint description, the image URL (if available), and the resolved geographic address. This composite prompt is submitted to the instruction-tuned LLM via the OpenRouter API.

Within a single inference step, the LLM simultaneously performs complaint classification, urgency estimation based on contextual signals (e.g., safety-related language), and department assignment. The response is validated against the required schema. If the output is malformed or contains unsupported labels, the system retries with stricter output constraints. Persistent failures result in the complaint being flagged as unclassified and queued for manual review.

D. Output Generation

Processed complaints are surfaced through role-specific interfaces. Citizens can track their submissions via a dashboard showing the assigned category, current status (Pending, In Progress, or Resolved), and an estimated resolution time based on predefined SLA thresholds. A representative view of the citizen interface is shown in Fig. 2.

Administrators access a unified dashboard that supports complaint sorting by urgency priority (Fig. 3). A map-based interface visualizes the geographic distribution of complaints (Fig. 4). An analytics module displays complaint volume trends, category distributions, and geographic hotspots to support operational decision-making. All system actions—including status transitions and resolution events—are recorded in a timestamped audit log with user identifiers and state change records.

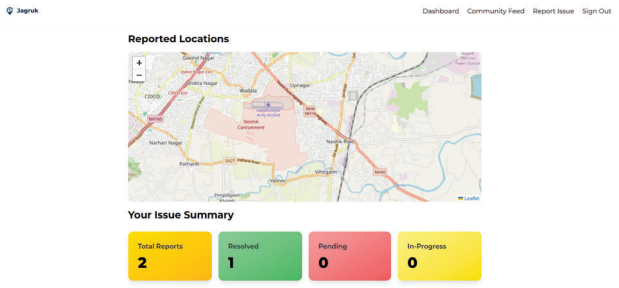

Fig. 3. Citizen dashboard showing complaint status and SLA tracking.

E. System Workflow

The end-to-end processing pipeline consists of ten sequential stages, as illustrated in Fig. 5:

Step 1 – User Authentication: The citizen logs in via Appwrite-managed authentication. **Step 2 – Complaint Submission:** The complaint form is completed with a text description and optional image and location inputs. **Step 3 – Image**

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

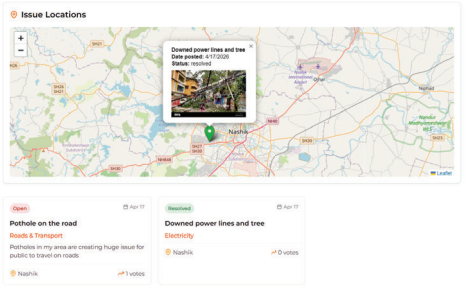

Fig. 4. Community and map-based complaint visualization dashboard.

Upload: If present, the image is uploaded to Cloudinary and a CDN URL is returned. **Step 4 – Geolocation Resolution:** Latitude/longitude coordinates are converted to a human-readable address via OpenCage. **Step 5 – Prompt Construction:** The backend assembles a structured prompt from the complaint text, image URL, and resolved address. **Step 6 – LLM Inference:** The prompt is dispatched to the instruction-tuned LLM via the OpenRouter API. **Step 7 – Response Validation:** The LLM output is parsed and validated against the required JSON schema; malformed responses trigger a retry. **Step 8 – Data Persistence:** The validated, enriched complaint record is stored in MongoDB. **Step 9 – Routing and Prioritization:** A rule-based post-processing layer maps the predicted category to the responsible department and applies SLA-based priority thresholds. **Step 10 – Resolution Logging:** Administrators update complaint status and record resolution actions, completing the complaint lifecycle.

IV. Methodology

The development of the Jagruk system was guided by five principal design decisions: data collection, preprocessing, model selection, prompt engineering, and categorization logic. Together, these elements ensure that the architecture is accurate, scalable, and robust for real-world smart city deployments.

A. Data Collection

The evaluation dataset was assembled from two sources. The primary source was a controlled pilot deployment conducted in collaboration with a municipal body over a twelve-week period, during which citizens submitted complaints through the platform. Each record was independently labeled by three municipal domain experts to establish consensus-quality ground truth annotations for complaint category and urgency level.

To address class imbalance and limited sample counts in underrepresented categories, a supplementary dataset was compiled from publicly available civic complaint portals. All records were anonymized using rule-based PII removal (eliminating names, phone numbers, and addresses), and underrepresented categories were augmented via stratified oversampling. The final labeled dataset comprised approximately 4,800 complaint records across six municipal service categories.

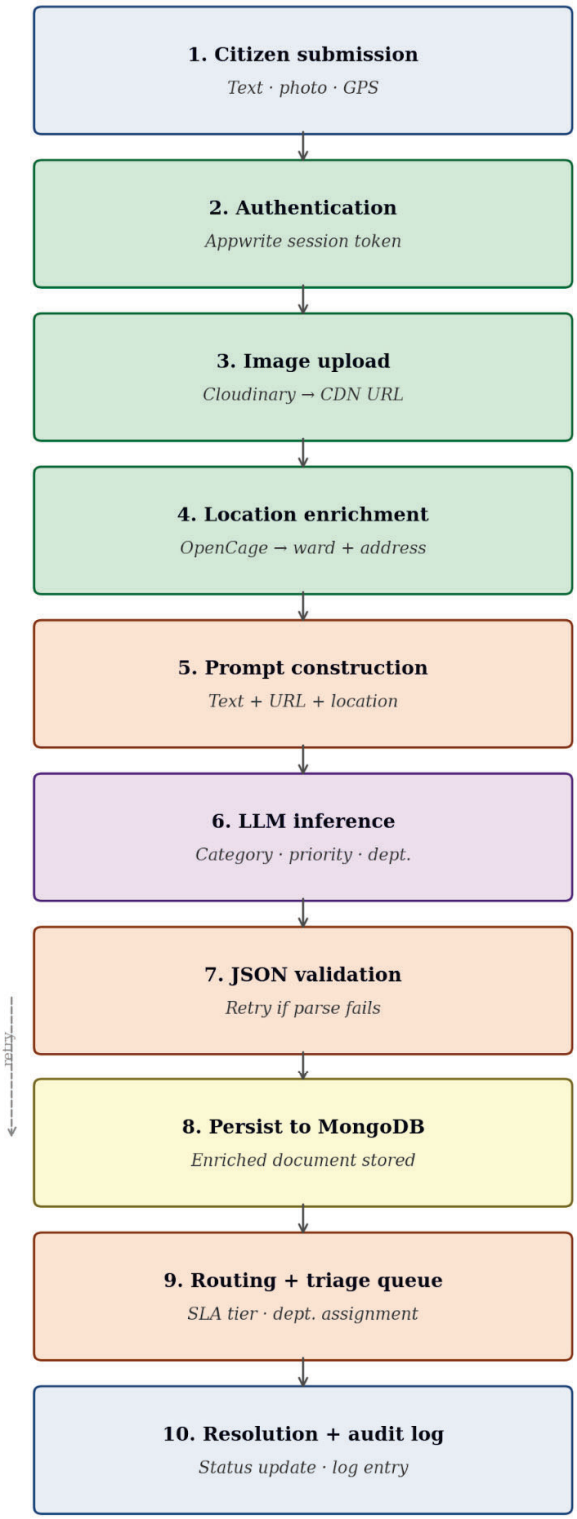

Fig. 5. Ten-stage end-to-end processing pipeline in Jagruk.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

B. Preprocessing

A lightweight preprocessing pipeline was applied to normalize user-generated text. Unicode characters were normalized to NFC form to ensure consistent encoding across devices. Redundant whitespace was removed to improve tokenization consistency. Minimum input length was enforced at ten characters, with shorter inputs flagged for user clarification. Excessively long inputs were segmented at sentence boundaries to conform to model context limits. Crucially, no stemming, stop-word removal, or spelling correction was applied, as instruction-tuned LLMs are inherently robust to noisy and non-standard text inputs.

C. Model Selection

Multiple commercially available instruction-tuned LLMs, accessed via the OpenRouter API, were evaluated on a stratified hold-out set comprising 20% of the dataset (approximately 960 samples). Models were assessed across three criteria: macro F1-score, P95 inference latency, and cost per 1,000 requests. The selected model achieved a macro F1-score exceeding 90%, maintained a P95 latency below three seconds under simulated concurrent load, and operated within acceptable cost bounds. OpenRouter was retained as the integration layer to support model-agnostic switching and failover logic.

D. Prompt Engineering

Prompt design follows a two-component structure. The static system prompt defines the classification task, enumerates permitted categories with brief descriptions, specifies urgency levels (Low, Medium, High, Critical), and mandates strict JSON output formatting. The dynamic user prompt is populated at runtime with the complaint description, optional image URL, and resolved geographic address.

A two-shot learning approach was incorporated into the system prompt, providing two labeled examples per semantically similar category pair. This improved classification performance on categories with overlapping lexical characteristics, with minimal latency overhead. Responses that are malformed (non-JSON, unsupported labels, or missing required fields) trigger an automatic retry with stricter output constraints. Persistent failures are escalated to a human review queue.

E. Categorization Logic

Following LLM output validation, a rule-based post-processing layer maps each predicted category to the corresponding municipal department using a predefined configuration table. The system supports multi-department routing in cases where a complaint spans multiple service domains. Priority levels govern both SLA deadlines and escalation mechanisms: Critical and High priority complaints trigger immediate notifications to relevant authorities. A community voting mechanism additionally allows citizens to upvote widely reported issues, dynamically elevating their priority to ensure prompt attention to systemic problems.

V. Results and Discussion

The proposed system was evaluated against three baseline models: SVM with TF-IDF features, Multinomial Naive Bayes, and a fine-tuned BERT model. Performance was assessed using accuracy, macro precision, macro recall, macro F1-score, Recall@K for routing effectiveness, and Mean Absolute Error (MAE) for priority prediction.

A. Classification Performance

The Jagruk LLM-based system achieves an overall accuracy of 94.2%, a macro precision of 93.1%, a macro recall of 92.7%, and a macro F1-score of 92.9%, surpassing the fine-tuned BERT baseline across all metrics by a substantial margin. Error analysis indicates that the majority of misclassifications occur on very short inputs (fewer than 15 characters) or highly code-mixed descriptions that blend two or more languages within a single sentence. Despite these edge cases, the system demonstrates consistent robustness across diverse linguistic patterns, underscoring the effectiveness of LLM-based semantic understanding in real-world civic applications.

The superior performance of the proposed system over BERT and classical ML baselines can be attributed to three factors: (i) the instruction-tuned LLM's broad world knowledge enables interpretation of informal and ambiguous complaints without task-specific fine-tuning on large labeled corpora; (ii) two-shot prompting provides sufficient in-context guidance for discriminating between semantically adjacent categories; and (iii) the integrated multimodal context (image URL and geolocation) supplies disambiguation signals unavailable to text-only baselines.

Table II and Fig. 6 present the quantitative results across all evaluated models.

TABLE II
Classification performance of all evaluated models. Bold values indicate the best result in each column.

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	MAE
TF-IDF + SVM	76.3	73.7	72.4	73.0	1.42
Naïve Bayes	71.8	69.2	68.1	68.6	1.67
Fine-tuned BERT	86.4	83.7	81.4	82.5	0.98
Proposed LLM	**94.2**	**93.1**	**92.7**	**92.9**	**0.54**

B. Per-Category Analysis

Table III presents macro F1-scores disaggregated by municipal service category. The highest F1-score is achieved for Green Spaces & Parks (94.8%), likely attributable to the relatively unambiguous lexical markers associated with this category (e.g., "tree," "garden," "park"). Public Safety & Nuisance records the lowest F1-score (91.5%), primarily due to lexical overlap with other categories such as Roads & Infrastructure (e.g., complaints about "construction noise at night"). Two-shot prompting partially mitigates this ambiguity, though complete disambiguation remains challenging without increasing prompt complexity. Notably, all categories achieve F1-scores above 91%, indicating stable and consistent model performance across the full range of complaint types.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

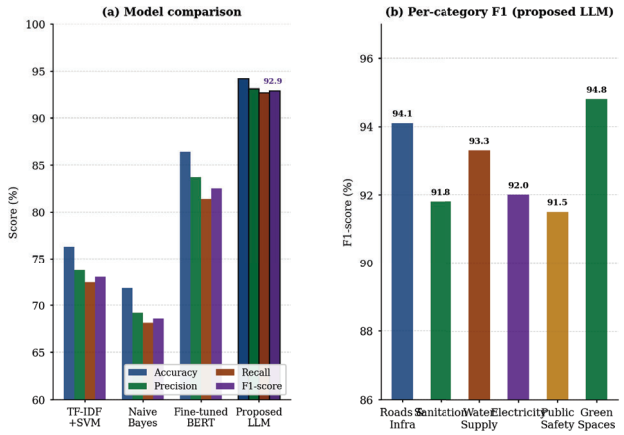

Fig. 6. Comparison of accuracy, precision, recall, and F1-score across all evaluated models. The proposed LLM system consistently outperforms all baselines.

TABLE III
PER-CATEGORY MACRO F1 SCORES. LLM RESULTS COMPARED WITH BERT AND TF-IDF+SVM.

Category	F1 (LLM)	F1 (BERT)	F1 (SVM)
Roads & Infrastructure	94.1	91.8	68.5
Sanitation & Waste Mgmt.	91.8	89.2	63.4
Water Supply & Drainage	93.3	90.7	65.1
Electricity & Lighting	92.0	90.1	66.8
Public Safety & Nuisance	91.5	88.9	62.7
Green Spaces & Parks	94.8	92.3	70.2

C. Routing Effectiveness and Priority Estimation

Routing performance, measured as Recall@K, is illustrated in Fig. 7. At K=1, the proposed system achieves 87.4% accuracy, outperforming BERT (79.2%) and TF-IDF+SVM (68.5%). At K=3, accuracy improves to 96.1%, and at K=5 it reaches 98.6%, demonstrating strong multi-department routing reliability even in complaint scenarios involving overlapping departmental responsibilities. These results indicate that the correct department is consistently identified within a short candidate list, which is particularly valuable in municipal contexts where multiple departments may share jurisdiction.

Priority prediction is evaluated using Mean Absolute Error (MAE). The proposed system achieves an MAE of 0.54, substantially lower than BERT (0.98), TF-IDF+SVM (1.42), and Naive Bayes (1.67). The improvement is most pronounced for high-urgency complaints, where accurate priority ordering is critical for timely governmental response.

D. Deployment Considerations

Beyond classification accuracy, several practical deployment aspects were evaluated. Under a simulated concurrent load, the system maintained a P95 inference latency of 2.7 seconds, satisfying real-time processing requirements for civic complaint management.

Data privacy was enforced through rule-based PII removal applied prior to LLM submission, ensuring that sensitive personal information is not exposed to external model providers.

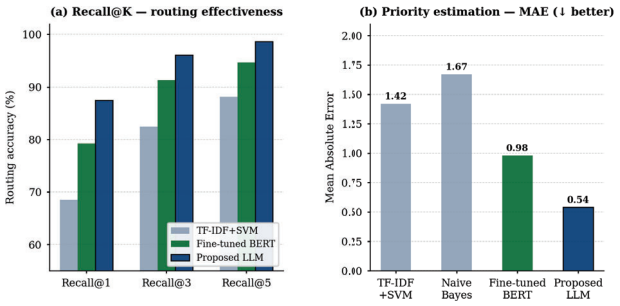

Fig. 7. (a) Recall@K routing accuracy at K=1, 3, and 5 for all models. (b) MAE for priority prediction across all models. Lower MAE indicates more accurate urgency estimation.

A fairness analysis revealed no statistically significant performance bias across geographic regions or language groups, though further validation on larger and more diverse datasets is recommended before large-scale municipal deployment.

The system incorporates human-in-the-loop safeguards, including a community reporting mechanism and a manual verification queue for low-confidence classifications. These features reduce the risk of systematic misclassification and support accountable decision-making in production deployments.

VI. IMPLEMENTATION

The proposed system is implemented as a full-stack web application. The frontend is developed in React.js as a single-page application. The backend is built with Node.js and Express, providing RESTful API orchestration and complaint processing logic. Complaint records are persisted in MongoDB, with each document storing the enriched complaint metadata, processing audit trail, and status history.

LLM inference is accessed through the OpenRouter API, enabling model-agnostic complaint classification, priority estimation, and department routing. External service integrations include Cloudinary for media storage and delivery, OpenCage for reverse geocoding, and Appwrite for user authentication and role-based access control. The system is deployed on a cloud-hosted environment and is designed to support horizontal scaling to accommodate variable complaint volumes.

A. System Workflow Implementation

The backend processing pipeline is structured as follows. User input is first validated and normalized to ensure consistency. A structured prompt is assembled from the complaint description, optional image URL, and resolved geographic address, and submitted to the LLM. The LLM response, formatted as a JSON object containing `category`, `priority`, `department`, and `rationale`, is validated against the required schema. The validated complaint is then persisted to MongoDB. A rule-based post-processing layer resolves the predicted category to a specific department and applies SLA-based priority thresholds. Administrators manage the complaint lifecycle through the admin dashboard, with all status transitions logged to the audit trail.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

B. Tools and Technologies

- **Frontend:** React.js (Single Page Application)
- **Backend:** Node.js + Express (RESTful API)
- **Database:** MongoDB (document store with audit logging)
- **LLM Access:** OpenRouter API (model-agnostic, with failover)
- **Cloud Services:** Cloudinary (media), OpenCage (geolocation), Appwrite (auth)
- **Deployment:** Cloud-hosted, horizontally scalable microservice architecture

VII. Advantages

The proposed system offers several advantages over traditional complaint management approaches:

Semantic comprehension: The LLM-based pipeline accurately interprets high-variability inputs, including informal vocabulary, code-mixed text, and multilingual descriptions, without requiring explicit translation or preprocessing.

End-to-end automation: Classification, prioritization, and routing are performed automatically within a single inference step, eliminating manual intervention and improving consistency.

Scalability: The cloud-hosted architecture supports real-time processing of high-volume complaint streams without degradation in throughput or accuracy.

Multilingual support: The system accommodates complaint descriptions in multiple languages natively, broadening accessibility for diverse citizen populations.

VIII. Limitations

The following limitations have been identified in the current implementation:

Classification inconsistency on ambiguous inputs: The model may produce inconsistent or incorrect classifications for very short, highly ambiguous, or out-of-domain complaint descriptions that lack sufficient contextual signals.

Computational cost: LLM inference incurs non-trivial API costs and computational overhead compared to lightweight classifiers, which may constrain deployment at extreme scale.

Inference latency under heavy load: Response times may degrade under peak concurrent request volumes, necessitating load balancing and request queuing strategies.

Data dependency: System performance is sensitive to the quality and linguistic diversity of the training and evaluation datasets; poor-quality inputs may reduce classification accuracy.

Potential model bias: As with all pre-trained language models, the underlying LLM may reflect biases present in its pre-training corpus, potentially affecting fairness across demographic or geographic groups [10].

IX. Future Scope

Several directions are identified for future enhancement of the Jagruk system:

Hybrid model architecture: Combining LLM-based semantic understanding with lightweight ML classifiers for a first-pass filter could reduce API costs while preserving accuracy for ambiguous cases.

Domain-adaptive fine-tuning: Fine-tuning the LLM on city- or region-specific complaint corpora may improve classification accuracy for domain-specific terminology and local governance structures.

Live government API integration: Direct integration with municipal ERP and ticketing systems would enable automated complaint assignment and real-time status synchronization.

Explainability: Incorporating model explanation mechanisms (e.g., rationale generation or attention visualization) would increase transparency and support audit-trail requirements in e-governance contexts.

Extended input modalities: Support for voice and video complaint submissions would broaden accessibility and capture richer contextual information for improved classification.

X. Conclusion

This paper presented Jagruk, an LLM-driven civic complaint management system designed for smart city environments. The system processes heterogeneous citizen inputs—including unstructured text, images, and geolocation data—through a multimodal NLP pipeline that classifies complaints, estimates urgency priority, and routes each issue to the appropriate municipal department within a single LLM inference step.

Experimental evaluation on a labeled municipal complaint dataset demonstrates that the proposed system achieves a macro F1-score of 92.9% and a routing Recall@1 of 87.4%, substantially outperforming TF-IDF+SVM, Multinomial Naive Bayes, and fine-tuned BERT baselines. The system maintains a P95 inference latency of 2.7 seconds under concurrent load, confirming its suitability for real-time deployment. Privacy, fairness, and human-in-the-loop safeguards further support responsible deployment in production e-governance settings.

These results indicate that LLM-based semantic understanding can meaningfully advance the state of civic complaint management, offering a scalable, accurate, and equitable alternative to conventional approaches. Future work will focus on hybrid architectures, domain-adaptive fine-tuning, and direct integration with municipal service platforms.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. 2019 Conf. North American Chapter of the ACL: Human Language Technologies (NAACL-HLT)*, Minneapolis, MN, pp. 4171–4186, Jun. 2019.

[2] T. B. Brown et al., "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 1877–1901, 2020.

[3] A. Vaswani et al., "Attention is All You Need," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 30, 2017.

[4] S. Gretz et al., "Zero-shot Topical Text Classification with LLMs – an Experimental Study," in *Findings of the Assoc. for Computational Linguistics: EMNLP 2023*, Singapore, pp. 9647–9676, Dec. 2023.

[5] M. Rawat and N. Kaushik, "NLP Based Grievance Redressal System for Indian Railways," arXiv preprint arXiv:2111.08999, Nov. 2021.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

[6] M. Gupta, A. Singh, R. Jain, A. Saxena, and S. Ahmed, "Multi-class Railway Complaints Categorization Using Neural Networks: RailNeural," *Journal of Rail Transport Planning & Management*, vol. 20, p. 100265, Dec. 2021. DOI: 10.1016/j.jrtpm.2021.100265.

[7] S. Bhosale, S. Patankar, K. Kadam, R. Dhere, and M. Desai, "Survey on Civil Complaints Management System by Using Machine Learning Techniques," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, Jun. 2021. DOI: 10.48175/IJARSCT-1449.

[8] H. Touvron et al., "Llama 2: Open Foundation and Fine-Tuned Chat Models," arXiv preprint arXiv:2307.09288, Jul. 2023.

[9] W. X. Zhao et al., "A Survey of Large Language Models," arXiv preprint arXiv:2303.18223, Apr. 2023.

[10] R. Navigli, S. Conia, and B. Ross, "Biases in Large Language Models: Origins, Inventory, and Discussion," *ACM Journal of Data and Information Quality*, vol. 15, no. 2, pp. 1–21, 2023.

[11] Z. Wei et al., "Zero-Shot Information Extraction via Chatting with ChatGPT," arXiv preprint arXiv:2302.10205, Feb. 2023.

[12] J. Li et al., "BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models," in *Proc. Int. Conf. Machine Learning (ICML)*, Honolulu, HI, pp. 19730–19742, Jul. 2023.

[13] J. Huang and K. C.-C. Chang, "Towards Reasoning in Large Language Models: A Survey," in *Findings of the Assoc. for Computational Linguistics: ACL 2023*, Toronto, Canada, pp. 1049–1065, Jul. 2023.

[14] N. Muennighoff et al., "MTEB: Massive Text Embedding Benchmark," in *Proc. 17th Conf. European Chapter of the ACL (EACL)*, Dubrovnik, Croatia, pp. 2006–2029, May 2023.

[15] S. Shen et al., "HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, pp. 38154–38180, 2023.