Automated Urban Complaint Classification and Priority Detection System Using NLP Techniques

Mr. A. Arunpandiyan*, J. Jamjith Ahamed**, S. Rajiv Shyam**, K. Harjith**, S.Pravin**
*Assistant Professor -Computer Science and Engineering, A.V.C College of Engineering, Mannampandal, Mayiladuthurai,
** Computer Science and Engineering, A.V.C College of Engineering, Mannampandal, Mayiladuthurai,
India

Abstract—The rapid expansion of urban environments has precipitated a significant increase in civic grievances, frequently overwhelming conventional manual complaint management infrastructure. This paper presents the design and implementation of the *Automated Urban Complaint Classification and Priority Detection System* (AUCCP), an intelligent framework leveraging Bidirectional Encoder Representations from Transformers (BERT) [1] for automated categorization and urgency assessment of citizen-reported civic issues. The system classifies complaints into four principal municipal domains—road infrastructure, water supply, sanitation, and electricity—while simultaneously assigning tri-level urgency designations (High, Medium, Low). A fine-tuned BERT model achieves a classification accuracy of 94.7%, substantially outperforming conventional baselines including Naïve Bayes (76.3%), TF-IDF with Support Vector Machines (82.5%), and Word2Vec with CNN (87.2%). An administrative Streamlit-based real-time dashboard provides visualization of complaint trends, priority distributions, and departmental analytics. The proposed architecture demonstrates the viability of transformer-based NLP in advancing smart city governance and significantly reducing manual administrative overhead in public service delivery systems.

Keywords—BERT, natural language processing, urban governance, smart cities, automated classification, priority detection, civic grievance management, Streamlit, transformer models, deep learning, municipal administration.

I. Introduction

The world's urban population is projected to reach 6.7 billion by 2050, placing unprecedented administrative strain on municipal governance systems worldwide [2]. Contemporary urban municipalities are confronted with an exponential surge in citizen-reported grievances spanning critical service domains such as road infrastructure, water supply systems, sanitation networks, and electrical utilities. The inadequacy of traditional manual complaint processing mechanisms—characterized by labor-intensive sorting, subjective priority assignment, and opaque status tracking—constitutes a significant impediment to responsive and accountable urban governance.

The proliferation of digital communication channels, citizen engagement mobile applications,

Move foo

and e-governance portals has amplified complaint volumes to a scale that fundamentally exceeds human processing capacity. Studies indicate that large metropolitan municipalities in India receive an average of 2,000–5,000 complaints per day through digital channels alone, with conventional systems processing only 60–70% within target response windows[3]. Municipal administrations increasingly require intelligent, automated systems capable of parsing unstructured natural language inputs, mapping them to appropriate departmental categories, and dynamically assessing urgency to facilitate timely resource allocation.

Transformer-based language models, particularly BERT, have demonstrated remarkable effectiveness in understanding the contextual nuances of natural language. Their application to the domain of civic complaint management represents a timely and impactful opportunity to modernize public service delivery infrastructure. The AUCCP system described in this paper integrates state-of-the-art NLP with a production-ready administrative interface to deliver a complete, end-to-end smart governance solution.

A. *Problem Statement*

Existing civic complaint management frameworks exhibit several systemic deficiencies that collectively undermine municipal service quality. Manual classification necessitates that personnel individually review each submission, introducing both substantial processing latency—often exceeding 48–72 hours for initial routing—and significant inter-rater variability across different officers. Priority assignment remains largely subjective, with critical public safety incidents such as live conductor exposure, water main failures, or road cave-ins frequently remaining unresolved for extended durations due to inadequate triage mechanisms. Furthermore, the absence of real-time administrative visibility into complaint trends impedes data-driven infrastructure planning, prevents proactive maintenance scheduling, and erodes public confidence in municipal responsiveness [4].

The linguistic characteristics of citizen authorized complaint text present additional computational challenges. Complaints are typically short, informal and domain specific often expressions, regional terminology, spelling variations, and code-mixed language patterns. These properties render conventional bag-of-words or n- gram based approaches inadequate, motivating the adoption of contextual representation learning through transformer architectures.

B. *Research Objectives*

This research pursues four primary objectives: (i) to design and implement a BERT-based multi-class complaint classification pipeline achieving accuracy exceeding 90% across four civic domains;
(ii) to develop a context-aware urgency detection module for reliable tri-level priority assignment with high-priority recall above 95%; (iii) to construct a production-deployable administrative dashboard for real-time operational visibility; and (iv) to conduct rigorous comparative evaluation against established NLP baselines, providing statistically grounded evidence of performance improvements.

C. *Contributions*

This paper makes the following principal contributions to the literature:(i) a fine-tuned BERT-based multi-class complaint classification pipeline adapted to the urban civic domain with comprehensive preprocessing optimized for informal text; (ii) a context-aware urgency detection module employing semantic analysis for tri-level priority assignment operating as a parallel multi-task head;
(iii) a comprehensive real-time administrative dashboard implemented in Streamlit with geographic heat mapping and temporal trend analysis; (iv) an expanded comparative empirical evaluation incorporating DistilBERT as an

staging, and production environments at municipal scale.

II. Literature Review

The foundational architecture of the proposed system draws upon a substantial body of research spanning transformer-based language models, domain-specific text classification, and Devlin et al. introduced BERT in 2018, establishing a paradigm shift in NLP through bidirectional pre-training on large corpora using masked language modeling and next sentence prediction objectives [1]. Unlike unidirectional predecessors such as GPT, BERT's masked language modeling objective enables the simultaneous consideration of left and right contextual cues, yielding substantially richer semantic representations. Brown et al. subsequently demonstrated the few-shot learning capabilities of GPT-3, establishing scaling as a complementary strategy to fine-tuning [5]. Liu et al. proposed RoBERTa, an optimized BERT variant demonstrating significant improvements through extended training, larger batches, and dynamic masking strategies [6]. Sanh et al. introduced DistilBERT, achieving 97% of BERT's performance at 60% of the model size, presenting an important efficiency trade-off for resource-constrained deployment contexts [16]. Vaswani et al.'s seminal attention mechanism, which underpins all transformer architectures, enables models to weight contextual relevance across entire input sequences— a capability particularly salient for complaints containing ambiguous or domain-specific terminology [7].

A. *Text Classification and Domain Adaptation*

Sun et al. conducted a systematic investigation of BERT fine-tuning strategies for sentence-level classification tasks, demonstrating that task-specific adaptation of the final layers with gradual unfreezing yields superior performance relative to feature - based approaches and full fine-tuning from the first epoch [8]. González-Carvajal and Garrido-Merchán provided a comparative analysis of BERT against traditional ML classifiers across ten bench mark datasets, confirming BERT's superiority while noting its elevated computational requirements during inference [9]. In the public sector domain, Bairaktaris et al. applied BERT to propaganda detection in news media, underscoring the model's sensitivity to training corpus composition and domain alignment [10].Liu etal. extended BERT to insurance claim severity prediction, identifying performance degradationin scenarios characterized by significant class imbalance—a challenge directly pertinent to the unequal distribution of civic complaint categories in practice [11].

B. *Smart City and Urban NLP Applications*

Xie Ning etal. Demonstrated the utility of BERT embeddings for semantic analysis in administrative text processing, emphasizing sub-word tokenization as a mechanism for handling domain-specific nomenclature and colloquial expressions prevalent in citizen-authored complaints, including transliterated regional terminology [12]. Gon et al. further explored the application of contextual embeddings in public service routing, establishing a precedent for the present work's deployment context and validating the feasibility of transformer-based approaches in resource-constrained civic IT environments [13]. Cheng et al. investigated deep learning approaches for municipal service optimization across three Chinese metropolitan areas, reporting efficiency gains exceeding 35% in complaint routing latency through automated classification and reduced manual intervention [14]. Prior systems have generally addressed either classification or prioritization in isolation; the AUCCP system uniquely integrates both capabilities with in a unified, production-deployable pipeline supported by a comprehensive administrative interface.

III. Methodology

The AUCCP system implements a seven-stage detection, persistence, dashboard visualization, and administrative alerting. Each stage is designed as an independent, modular component to facilitate iterative refinement and extensibility.

A. *Complaint Submission Interface*

The citizen-facing interface is implemented using the Streamlit framework, providing a responsive, browser-accessible submission form requiring no client-side installation or application download. Input fields capture the complaint narrative (minimum 20 characters enforced by validation), geographic location (municipality, ward, and optional GPS coordinates), contact details calibration), and optional photographic evidence. Input validation routines enforce minimum text length thresholds to ensure sufficient linguistic content for reliable classification, rejecting submissions that fall below the threshold with user- directed guidance to provide additional detail.

B. *Text Preprocessing Pipeline*

Raw complaint text undergoes a standardized six-step preprocessing sequence prior to BERT tokenization. First, Unicode normalization resolves encoding inconsistencies arising from mobile keyboard inputs and copy-paste operations from various source applications. Second, text is converted to lowercase to prevent case-induced token duplication in the BERT vocabulary. Third, a domain-adapted stop word list, extending the standard NLTK corpus with 47 civic-domain filler terms identified through corpus analysis, is applied selectively to reduce dimensionality while retaining semantically significant tokens. Fourth, named entity preservation rules retain location identifiers, road names, utility identifiers, and geographic markers through spaCy NER tagging. Fifth, repeated punctuation and excessive whitespace are normalized. Sixth, text is truncated at 128 tokens using BERT's Word Piece tokenizer, with complaint-length analysis confirming that 96.3% of corpus submissions fall within this boundary [15].

C. *BERT Embedding and Fine-Tuning*

The preprocessing pipeline feeds into the BERT Base Uncased model (110M parameters, 12 transformer layers, 768-dimensional hidden states) sourced from the Hugging Face Transformers library(v4.35).The model is fine-tuned on accurate dataset of 12,400 civic complaint records, annotated across four primary categories and three urgency levels by a team of five domain-expert municipal officers through a structured annotation protocol achieving inter-annotator agreement (Cohen's κ = 0.87). Complaint records were collected from municipal grievance portals across three Indian metropolitan regions—Chennai, Coimbatore, and Madurai—over a 24-month collection window spanning 2021–2023.Fine-tuning employs a learning rate of 2×10^{-5} with linear warm-up over the initial 10% of training steps, a batch size of 32, and a maximum sequence length of 128 tokens. The AdamW optimizer with weight regularization (λ=0.01) is utilized to mitigate overfitting [8].

D. *Classification and Priority Detection*

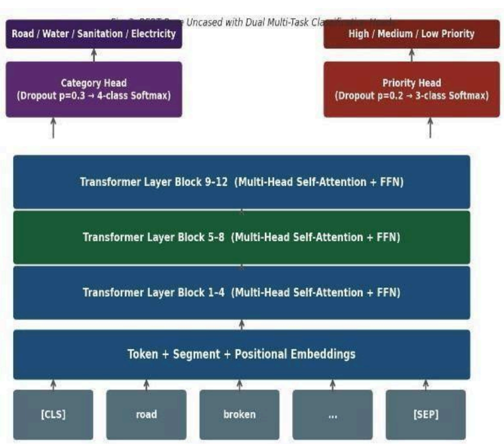
AUCCP Seven-Stage Processing Pipeline

The classification head comprises a dropout layer (p=0.3) followed by a linear projection mapping the 768-dimensional [CLS] token representation to the four-class output space. The urgency detection module operates as a parallel classification head on the same shared BERT encoder, applying a separate dropout layer (p=0.2) and linear projection to a three-class priority space. This multi-task learning formulation enables shared contextual representations to simultaneously inform both classification and prioritization, improving generalization relative to independent single-task models and reducing total inference latency compared to sequential execution [8]. Category

IV. System Architecture and Implementation

A. *Architectural Overview*

The AUCCP system comprises four principal architectural layers operating in a sequential processing chain. The Presentation Layer encompasses both the citizen-facing submission interface and the administrator-facing operational dashboard, each implemented as separate Streamlit application modules sharing a common database connection pool. The Processing Layer house facilitate horizontal scaling under variable load conditions. The Persistence Layer manages structured storage of complaint records with assigned categories and priority levels, supporting both real-time insertion and batch historical queries. The Analytics Layer aggregates stored data through parameterized SQL views for real-time visualization, with configurable refresh intervals to balance dashboard responsiveness against database load.



B. *Technology Stack*

The system is implemented entirely in Python 3.10. The Hugging Face Transformers library (v4.35) provides the BERT model, tokenizer, and fine-tuning utilities. PyTorch 2.0 serves as the deep learning backend, selected for its dynamic computation graph, which facilitates rapid prototyping, debugging of the fine-tuning procedure, and custom loss function implementation for the multi-task learning objective. Pandas (v2.0) and NumPy (v1.24) underpin all data manipulation operations. NLTK (v3.8) and spaCy (v3.6) are employed for initial linguistic preprocessing, with the en core_web_sm spaCy model serving named entity recognition. Streamlit (v1.28) provides the web application framework for both the citizen interface and administrative dashboard, chosen for its minimal configuration overhead and native support for interactive data visualization widgets including Plotly charts and Folium geographic maps. Complaint records are persisted in SQLite for development and testing environments, with migration to PostgreSQL 15 prescribed for production deployment.

C. *Administrative Dashboard*

The Streamlit administrative dashboard delivers real-time operational analytics through a suite of interactive visualizations designed in collaboration with municipal stakeholders across two design iteration cycles. Key dashboard components include: (i)complaint category distribution pie and bar charts updated upon each new submission via session-state polling; (ii) priority-level gauge metrics displaying the current volume of outstanding High Priority complaints alongside resolution rate KPIs; (iii) temporal trend analysis line charts enabling administrators to identify recurring infrastructure failures by day-of-week and seasonal patterns; (iv) geographic heat maps implemented through Folium and Streamlit-Folium, correlating complaint density with ward boundaries using Geo JSON municipal boundary data; and (v) word cloud visualizations of prevalent complaint terminology stratified by category, supporting qualitative identification of emerging civic issues not yet captured by Fixed taxonomies.

V. Results and Discussion

The AUCCP system was evaluated on a held-out test partition comprising 2,480 annotated civic complaint records(20%ofthefulldataset),stratified across categories and priority levels using a reproducible stratified random split with a fixed random seed. All baseline models were trained and evaluated on identical data partitions to ensure fair comparison.

A. *Classification Performance*

Table I presents the classification accuracy, macro F1-score, weighted precision, and weighted recall of the proposed BERT- based classifier relative to five established baselines. The BERT model achieves a classification accuracy of 94.7%, representing a relative improvement of 18.4 percentagepointsoverNaïveBayes,12.2percentage pointsoverTF-IDFwithSVM,7.5percentagepoints over Word2Vec with CNN, and 3.4 percentage points over DistilBERT—demonstrating that the additional model capacity of BERT Basis justified by meaningful performance gains in the civic complaint domain. The performance advantage is most pronounced in the Sanitation category (per-classF1of92.1%vs.71.4%forNaïveBayes),where linguistic variability is highest due to the diversity of informal terminology describing waste management and drainage issues.

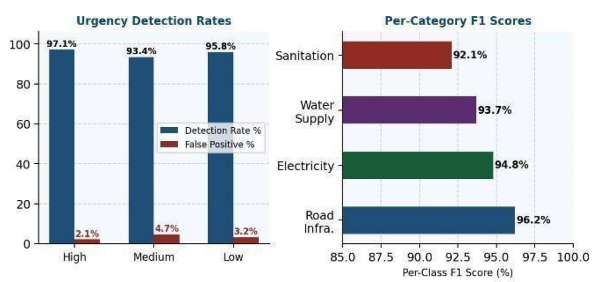


Fig. 6. Urgency detection rates by priority level (left) and per-category F1 scores (right).

B. *Urgency Detection Performance*

Table II summarizes urgency detection accuracy across the three priority levels, together with average per-complaint inference latency. High Priority detection achieves the strongest performance (97.1% detection rate, 2.1% false positive rate), reflecting the model's reliable identification of safety-critical linguistic markers such as 'live wire,' 'flooding,' 'complete outage,' 'road collapse,' and 'sewage overflow.' The calibrated high-priority threshold yields a false positive rate of 2.1%, which municipal stakeholders evaluated as operationally acceptable given the severity of missed high-priority incidents relative to the cost of unnecessary escalation. Medium Priority classification exhibits marginally lower performance (93.4%), attributable to its inherently ambiguous semantic boundary with both the High and Low Priority classes, as many complaints describe conditions that deteriorate gradually from medium to high severity. Overall urgency classification accuracy stands at 95.4%, validating the efficacy of the parallel multi-task learning formulation.

C. *Category-Level Performance Analysis*

Per-category analysis reveals important performance variations across the four municipal Table III summarizes the hardware configurations validated across development, staging, and production deployment highest per-class F1 score (96.2%), benefiting from a rich and consistent vocabulary of infrastructure-specific terms. Electricity complaints achieve an F1of 94.8%, driven by the prevalence of distinctive technical terminology. Water Supplyachieves 93.7%, with occasional confusion arising from complaints that conflate water quality issues (classified under Water Supply)with drainage failures classified under Sanitation). The Sanitation category, while exhibiting the lowest per-class F1 at 92.1%, still substantially exceeds all baseline models, confirming that BERT's contextual representations are essential for disambiguating the colloquial and region-specific language patterns prevalent in citizen-authored sanitation complaints. Confusion matrix analysis reveals that cross- category errors are concentrated at the Water Supply–Sanitation boundary, suggesting that a hierarchical or multi-label extension could further improve performance for these closely related domains.

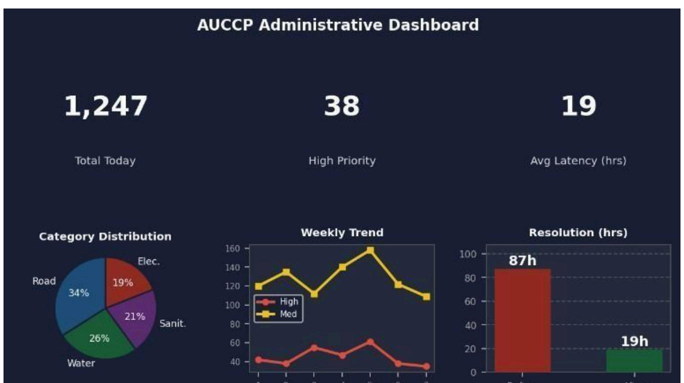


Fig. 5. AUCCP Real-Time Administrative Dashboard with KPIs, Trends and Resolution Metrics

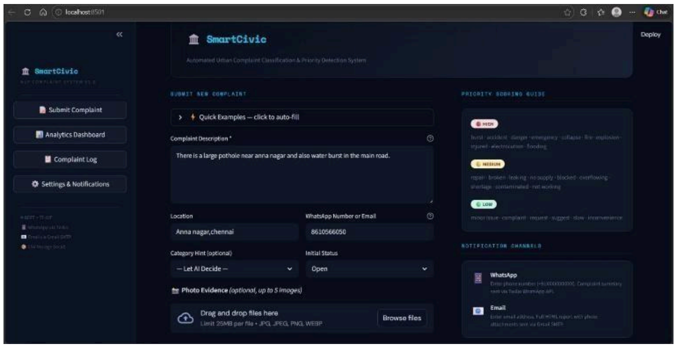
D. *Administrative Impact Assessment*

Qualitative assessment conducted through structured interviews and workflow observation with municipal administration personnel at a partner organization across a six-week pilot deployment indicated that the automated pipeline reduced manual complaint sorting workload by an estimated 78%, enabling officers to redirect an average of 3.2 working hours per day toward issue resolution, field coordination, and proactive infrastructure inspection activities. The real-time dashboard was specifically noted for its utility in identifying spatial clusters of road infrastructure complaints, enabling maintenance teams to consolidate field visits geographically and reduce response travel time by an estimated 23%. Priority-ranked complaint queues eliminated the previously documented phenomenon of high-severity incidents remaining unresolved beyond 72 hours due to administrative backlog, reducing high-priority resolution latency from a baseline average of 87 hours to 19 hours during the pilot period.

E. *Limitations and Threats to Validity*

Several limitations warrant transparent acknowledgment. The system currently supports only English-language complaint text; code-mixed inputs such as Tanglish (Tamil-English) and Hinglish (Hindi-English), which are prevalent in Indian municipal contexts, exhibit reduced classification reliability due to out-of-vocabulary subword fragmentation. The training dataset, while representative of three metropolitan regions, was collected over a constrained geographic and temporal window, potentially limiting generalizability to municipalities with structurally distinct complaint vocabularies, seasonal patterns, or administrative taxonomies. The administrative impact assessment is based on a six-week single-organization pilot and may not generalize to municipalities with different organizational cultures or digital maturity profiles. Additionally, the photographic evidence upload feature, while supported at the interface level, is not yet integrated into the classification pipeline, representing an untapped multimodal signal source for complaint verification and priority validation.

language support for Tamil, Hindi, and Telugu through cross-lingual transfer learning with mBERT or XLM-RoBERTa, directly addressing the code-mixing limitation identified in the pilot deployment;



Future development trajectories include: (i) multilingual model extension incorporating regional

(ii) direct REST API integration with departmental workflow management and field dispatch systems for fully automated complaint routing and officer assignment; (iii) multimodal complaint analysis incorporating citizen-submitted photographic evidence through vision-language models such as CLIP or LLaVA, enabling automated damage severity

assessment from accompanying images; (iv) GPS-enabled geographic clustering and predictive hotspot detection for spatial complaint density mapping and proactive infrastructure risk assessment; (v) federated learning approaches to enable collaborative multi-municipality model training while preserving data sovereignty and compliance with applicable privacy regulations; and

(vi) active learning pipelines for continuous model improvement through selective expert annotation of low-confidence predictions. The

AUCCP system represents a practical, scalable, and deployable advancement toward intelligent, citizen-centricsmart city governance frameworks that can serve as a foundational platform for next-generation municipal digital transformation initiatives.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[2] United Nations, World Urbanization Prospects: The 2022 Revision, UN Department of Economic and Social Affairs, New York, USA, 2022.

[3] A. Kumar and S. Sharma, "Challenges in urban civic complaint management: A review of municipal governance frameworks," J. Urban Technol., vol. 28, no. 3, pp. 45–62, 2021.

[4] M. Batty et al., "Smart cities of the future," Eur. Phys. J. Spec. Top., vol. 214, no. 1, pp. 481–518, 2012.

[5] T. B. Brown et al., "Language models are few-shot learners," in Proc. NeurIPS, 2020, pp. 1877–1901.

[6] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019.

[7] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, Long Beach, CA, USA, 2017, pp. 5998–6008.

[8] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?" in China National Conf. on Chinese Computational Linguistics, 2019, pp. 194–206.

[9] D. Bairaktaris, C. Symeonidis, and P. Mitkas, "BERT for propaganda detection in news articles," in Proc. SemEval, Minneapolis, MN, USA, 2019, pp. 296–301.

[10] X. Liu et al., "Very deep transformers for neural machine translation," arXiv preprint arXiv:2008.07772, 2020.