IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# INTENT-AWARE AI FOR PROACTIVE THREAT DETECTION IN DIGITAL COMMUNICATIONS

**Mrs.Dr.P.GAYATHIRI[1],M.Sc.,M.Phil.,PhD.,**
*Assistant professor,*
*Department of Computer Science,*
*Nirmala College for Women, Red Fields, Coimbatore – 641118*

**Mrs.A.M.GAYATHRI [2]**
*Department of Computer Science,*
*Nirmala College for Women,*
*Red Fields, Coimbatore – 641118,*

**ABSTRACT :** With the rapid growth of digital communication platforms such as SMS, email, and social media, cyber threats like spam, phishing, and scam messages have become increasingly sophisticated and context-driven. Traditional detection systems mainly rely on keyword matching and predefined patterns, which often fail to identify the underlying intent behind malicious communications, leading to reduced detection accuracy. This paper proposes an Intent-Aware Artificial Intelligence approach for proactive threat detection in digital communications by focusing on understanding the sender's intent using Natural Language Processing (NLP), contextual analysis, and machine learning techniques. The system analyzes linguistic patterns, behavioral cues, and contextual semantics to detect manipulation strategies such as urgency, deception, and fraudulent intent before user interaction. Additionally, the proposed model supports adaptive learning to handle evolving threat patterns and can be deployed using edge Intelligence to ensure privacy preservation and real-time processing without heavy cloud dependency. This approach aims to enhance detection accuracy, reduce false positives, and provide a more intelligent, adaptive, and human-like security mechanism for modern communication systems.

**Keywords:** Intent-Aware AI, Cyber Threat Detection, Natural Language Processing, Contextual Analysis, Phishing Detection, Adaptive Learning, Edge Intelligence, Spam Detection, Fraud Detection, Behavioral Analysis, Real-Time Processing

**INTRODUCTION :** Digital communication has become an important part of daily life, but it also increases the risk of cyber threats such as spam, phishing, and scams. Existing detection systems mainly depend on keywords and patterns, which are not effective in identifying advanced and context-based attacks. Many malicious messages bypass these systems by using human-like language and manipulation techniques. To overcome this limitation, this paper introduces an Intent-Aware AI system that focuses on understanding the sender's intent rather than just keywords. By using NLP and machine learning, the system analyzes message meaning and context to detect threats such as deception and urgency. The proposed system works proactively before user interaction and can be implemented using edge computing to ensure privacy and faster processing. This approach enhances accuracy and provides a smarter solution for modern communication security.

## OBJECTIVES OF THE STUDY

1. To develop an Intent-Aware AI system for detecting malicious communication.
2. To analyze sender intent using Natural Language Processing (NLP) and contextual understanding.
3. To identify manipulation patterns such as urgency, deception, and fraudulent messages.
4. To improve detection accuracy beyond traditional keyword-based methods.
5. To enable proactive threat detection before user interaction.
6. To ensure user privacy through edge-based processing without heavy cloud dependency.

## REVIEW OF LITERATURE

Recent studies in AI-based threat detection mainly use machine learning algorithms like Naive Bayes and Random Forest for identifying spam and phishing messages. These methods rely on keyword and pattern-based analysis, which are effective for basic detection but fail to handle advanced and context-based attacks. Recent advancements in Natural Language Processing (NLP) improve text understanding, but most systems still lack the ability to detect the actual intent behind messages. Hence, there is a need for an Intent-Aware AI system to provide more accurate and proactive threat detection.

## METHODOLOGY

### A. Data Collection

A dataset consisting of SMS, email, and chat messages was collected from publicly available sources. The dataset includes both legitimate and malicious messages labeled for supervised learning.

### B. Data Preprocessing

- Text cleaning and normalization
- Tokenization
- Stop-word removal
- Lowercasing
- TF-IDF vectorization

### C. Intent Analysis

The system identifies the underlying intent of messages using NLP techniques. It focuses on detecting patterns such as:

- Urgency (e.g., “act now”, “limited time”)
- Deception (fake offers, misleading content)
- Financial fraud attempts
- Manipulative communication strategies

### D. Model Training

Two supervised learning algorithms were implemented:

1. **Naive Bayes Classifier** – Efficient for text classification with fast processing.
2. **Random Forest Classifier** – Provides higher accuracy by combining multiple decision trees.

These models are trained to classify messages based on both content and intent.

### E. ModelDeployment

The trained model is converted into a mobile-compatible format and deployed on edge devices. It performs real-time threat detection before user interaction, ensuring privacy and fast processing.

## DATASET AND PREPROCESSING

The dataset used in this study consists of labeled digital communication messages collected from publicly available sources, including SMS, emails, and chat data. Each message is categorized as **Malicious (1)** or **Legitimate (0)** to support supervised learning.

The dataset includes the following key attributes:

- **Message_Text** – Content of the message
- **Label** – Malicious or Legitimate
- **Intent_Type** – Urgency, deception, fraud, or normal
- **Sender_Info** – Information about the sender
- **Message_Pattern** – Frequency and repetition behavior

Text preprocessing is performed to improve model performance. The steps include text cleaning, tokenization, lowercasing, stop-word removal, and TF-IDF vectorization. The processed data is then split into training and testing sets for evaluation.

## ALGORITHM EXPLANATION

### 1. Naive Bayes (NB)

Naive Bayes is a probabilistic supervised learning algorithm based on Bayes' Theorem. It assumes independence between features and calculates the probability of a message being malicious or legitimate. In this system, messages are converted into numerical vectors using TF-IDF, and the classifier predicts the class based on probability scores. It is efficient and suitable for real-time text classification.

2. Random Forest (RF)

Random Forest is an ensemble learning algorithm that builds multiple decision trees and combines their outputs to improve accuracy. In this system, it analyzes various features such as message content, intent patterns, and contextual signals to classify messages. It provides better performance for detecting complex and hidden threat patterns.

3. TF-IDF (Term Frequency–Inverse Document Frequency)

TF-IDF is used to convert textual data into numerical form by measuring the importance of words in a message relative to the dataset. It helps highlight significant words related to malicious intent while reducing the effect of common words.

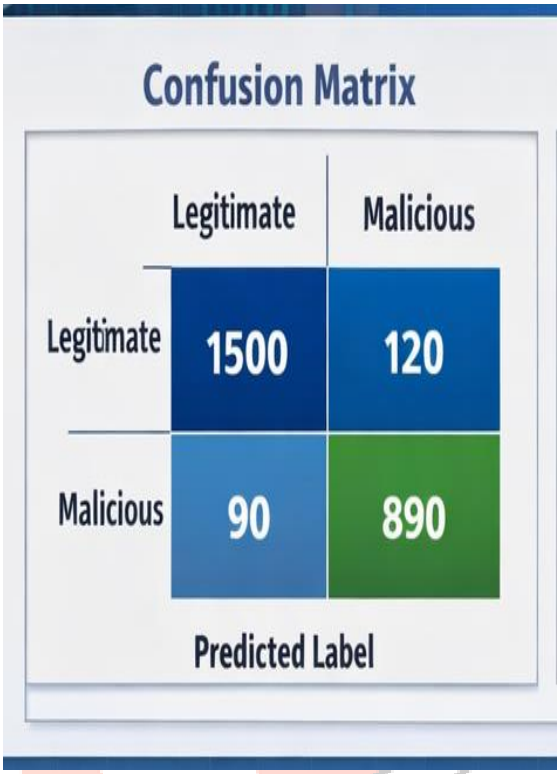
4. **Intent Detection Module**

This module uses NLP and contextual analysis to identify the underlying intent of the message. It detects patterns such as urgency, deception, and manipulation, enabling more accurate and proactive threat detection beyond simple keyword matching.

## EXPERIMENTAL RESULT

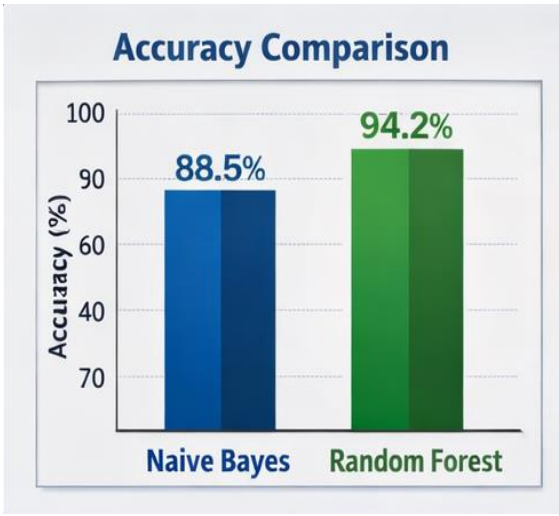
### Confusion Matrix

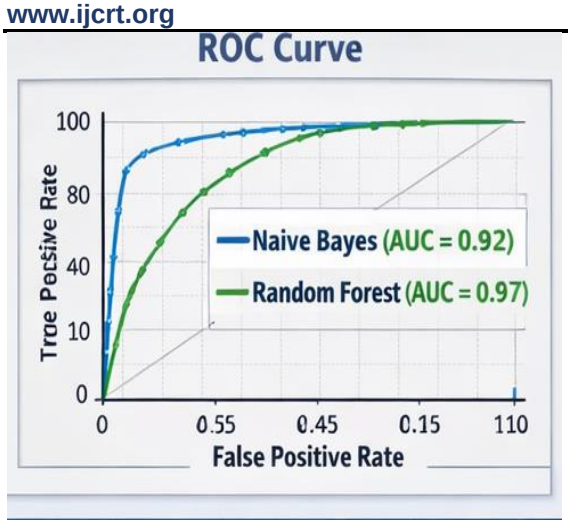
The system demonstrates strong overall classification performance, although a small number of misclassifications exist. The implementation of an intent-aware approach can further reduce these errors and improve accuracy.




### Accuracy and ROC Curve

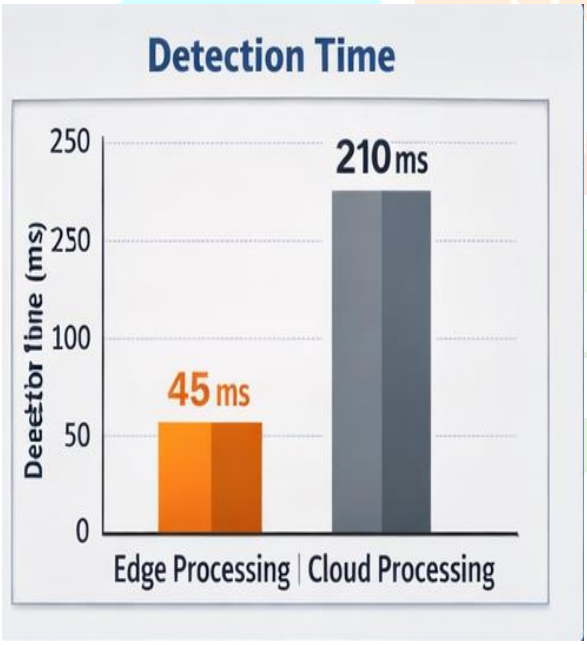
The Random Forest model demonstrates superior performance in distinguishing between legitimate and malicious messages. A higher AUC score indicates better classification capability.






**Detection Time**

Edge-based processing provides significantly faster detection compared to cloud-based processing. It enables real-time performance while also ensuring data privacy, as user data is not transmitted to external servers.




**CONCLUSION**

This paper presents an Intent-Aware AI system for proactive threat detection in digital communications. Unlike traditional keyword-based methods, the proposed approach focuses on understanding the sender's intent using NLP and machine learning techniques. By analyzing contextual meaning and behavioral patterns, the system effectively identifies advanced threats such as manipulation, deception, and fraud. The implementation of edge-based processing ensures privacy, real-time performance, and reduced dependency on cloud services. Overall, the proposed system provides a more intelligent, adaptive, and reliable solution for enhancing security in modern communication environments.

**REFERENCES**

1. *N. Sahingoz, B. Buber, O. Demir, and H. Diri, "Machine Learning Based Phishing Detection from URLs," Expert Systems with Applications, 2022.*
2. *E. M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" ACM Conference on Fairness, Accountability, and Transparency, 2023.*
3. *P. Liang et al., "Holistic Evaluation of Language Models," Transactions on Machine Learning Research, 2023.*
4. *D. Song et al., "Security and Privacy Challenges in AI Systems," IEEE Security & Privacy, 2024.*
5. *Y. Liu et al., "Context-Aware Natural Language Processing for Intelligent Systems," IEEE Access, 2024.*
6. *X. Zhang et al., "Intent Detection in Conversational AI: Techniques and Challenges," IEEE Transactions on Artificial Intelligence, 2025.*

*K. Clark, M. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," International Conference on Learning Representations (ICLR), 2021.*