International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

Ai Powered Document Forgery Detection and Intelligent Verification System using Machine Learning

Aavani Madhu
Student, Department of BCA, St Francis College

Kasinath KS
Student Department Of BCA, St Francis College,

Dr Geetha. S
Assistant Professor Senior St Francis College
Email Id: ask2geetha@gmail.com

Abstract - In fields such as education, banking, healthcare, and government, document forgery which leads to financial loss and insecurity is a huge problem and traditional means of verification are ineffective. This project proposes an AI Powered Document Forgery Detection and Intelligent Verification System using Machine Learning to automatically analyse and verify documents against forgeries. Users will first be required to login then upload the relevant digital document images (PDF, JPG, JPEG, PNG). Documents will be pre-processed into a high-quality image format ready for processing. Features such as text will be extracted using OCR. Text forgery, image forgery, copy-move forgery, irregular font and signature forgery will be detected using Machine Learning and Computer Vision algorithms. These detected features are then used to predict the documents as Genuine/Authentic and Forged/Tampered. The prediction scores will be highlighted as well, indicating the level of authenticity or forgery. A document fingerprint will be generated using the SHA-256 hashing algorithm and stored in the database together with the prediction results and owner details. By comparing hashes of two given documents, it will be possible to determine whether a document has been altered by checking the hashes of the two documents against each other. The application will generate downloadable PDF reports containing the document details, extracted text, and prediction results including confidence level for users to use as a reference. This system will provide an intelligent solution for document analysis and verification and will offer a wide range of security-related applications including document verification and fingerprinting using OCR, Machine Learning and Computer Vision, and hashing algorithms.

Keywords: Document Forgery Detection, Machine Learning, Optical Character Recognition (OCR), Computer Vision, SHA-256 Hashing

1. INTRODUCTION

With the rapid growth of digital technology comes the growth of creation, storage and exchange of digital documents. Documents are widely used in education, banking, healthcare, government and other sectors to store personal, financial, academic, and official information. The shift from using physical documents to digital documents has made information easier to access, store and share. This transition has, however, come at the expense of increased risk of document manipulation and forgery. With the availability of image-editing and document-editing tools, text, images, signatures, numbers and other document elements can be manipulated without being immediately noticed.

Document forgery is therefore an important security concern since a forged document can be used to provide false information or lead to financial losses or affect the reliability of critical records. In educational institutions, forged certificates and documents may be used to falsely represent qualifications. In banking and finance, altered documents may lead to fraudulent transactions. Likewise, manipulated healthcare and government documents may affect critical decision-making.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

The use of digital documentation has, in many areas, necessitated a need to verify the truthfulness and integrity of such documents. Traditional document verification methods generally involve manual inspection, comparing with original documents and verifying with authorized sources. Such methods can be helpful, but can be time consuming and require a significant amount of human effort especially if a large number of documents need to be analysed. Manual examination might also be futile, particularly if the changes are subtle, or if digital editing methods are employed to make a tampered document appear genuine. An automated approach that can assist with analysing digital documents and identifying possible signs of manipulation is therefore required.

Several technologies have been developed for different aspects of document analysing and forgery detection. OCR is applied to extract text from document images and convert it to machine-readable information. Image processing and Computer Vision techniques can be applied to analyse the visual features of documents and detect potential anomalies. Machine Learning techniques can be applied further to recognise patterns and classify documents based on extracted features. Another way to verify the integrity of digital documents is through cryptographic hashing. This method produces a unique hash value, which can be used as a digital fingerprint. The SHA-256 hashing algorithm can be used to generate such a print. Then, on the basis of this print, the verification of each document version against it can be performed to check if their digital content has been altered.

It is essential to note, however, that each of the aforementioned techniques is used for solving different aspects of the document verification task. Therefore, OCR is used to extract textual data, while Computer Vision and image processing algorithms are used to analyse the image. In turn, Machine Learning can be used to classify the documents, while hashing is used to verify if the digital content of the document has been altered. Thus, the document verification system based on multiple methods can be more effective than the one relying on one method only. A need therefore exists to develop an integrated system that can analyse both textual and visual characteristics while offering an additional mechanism for document integrity checking.

The objective of the current work is to develop an AI Powered Document Forgery Detection and Intelligent Verification System using Machine Learning. The proposed system enables users to securely log in and upload digital documents in PDF, JPG, JPEG and PNG formats. The documents uploaded by the user are pre-processed into a suitable image format before analysis. OCR is used to extract textual information, while Machine Learning and Computer Vision techniques are applied to analyse possible text forgery, image forgery, copy-move forgery, irregular font characteristics and signature forgery. The system classifies the document as Genuine/Authentic or Forged/Tampered, based on the analysed features, and provides a prediction score indicating the confidence level of the classification.

In addition to forgery detection, the proposed system generates a SHA-256 digital fingerprint for the uploaded document. The fingerprint can be stored along with the prediction result and relevant document information. Where two versions of a document are available, their hash values can be compared to determine whether the documents' digital contents are identical or have been altered. The system also generates a downloadable PDF report that includes the details of the document, the extracted text and prediction results including the confidence level.

The main contribution of the current work is the integration of OCR, Machine Learning, Computer Vision and SHA-256 hashing techniques in a single document verification workflow. Instead of relying on manual examination or a single verification technique, the proposed system combines document content analysis, visual forgery detection, authenticity classification and digital fingerprinting to form an intelligent solution for helping users identify potentially forged or tampered documents while storing the verification results.

2. LITERATURE REVIEW / RELATED WORK

The increasing need for the usage of digital certificates, several bills, receipts, academic statements, and other significant documents has raised the demand for trustworthy and automated verification techniques. Traditional verification techniques involve a manual examination and comparison of original documents and visual inspection. These techniques

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

can be extremely time-consuming or sometimes fail to detect even minor alterations that can be done using modern-day document and image-editing software. Researchers have been investigating the use of machine learning, Optical Character Recognition (OCR), computer vision, and cryptographic techniques for detecting and verifying forged documents.

Sirajudeen et al. [1] proposed an automatic document forgery detection system for such types of identity cards, certificates, passports, and legal documents. The study noted that the described process would create a substantial volume of administrative work, requiring an extended period of manual verification. The researchers utilized Convolutional Neural Networks (CNN), Optical Character Recognition (OCR), and Local Binary Pattern (LBP) algorithms to analyse the text information and the region-based features of the documents in question. This study serves as an excellent example of using text extraction with the aid of machine learning and how it can improve the document verification process. However, the approach mainly focuses on detection and does not provide a verification workflow involving document fingerprinting and comparison of document versions.

Sun, Ni, and Zhao [2] developed the Multi-Level Features Attention Network (MFAN) for detecting fake certificate images. Unlike many previous forensic studies that focused on natural images, this research considered certificate documents with different types and sizes of tampered regions. The method used an encoder-decoder architecture, the Atrous Spatial Pyramid Pooling (ASPP), low-level feature fusion, and channel attention to identify manipulation traces. The study covered operations such as splicing, copy-move, and object removal, but the method mainly concentrates on image-level detection without incorporating OCR-based textual analysis or cryptographic document integrity verification.

Boonkrong [3] proposed a document forgery detection system for academic documents and introduced Hashing as an integral mechanism for verifying document integrity. The article suggests utilizing a cryptographic hash to confirm if digital academic documents were altered and compares the effectiveness of hashing algorithm to the use of digital signatures, convolutional neural networks, and blockchain. The result of the research proves that using a cryptographic hash function would efficiently detect changes and rapidly authenticate the academic papers. The current study directly relates to the proposed system because it underlines the notion of implementing a unique tag to identify if documents were manipulated. However, the article does not thoroughly explain how someone could utilize a hash to indicate the particular location or type of forgery and falsified parts such as text, image, or signature.

Research has been conducted on financial and transactional documents like bills and receipts. Chakraborty [4] proposed an intelligent forgery detection approach for retail bills and receipts using a CNN combined with an optimization algorithm and texture-based feature extraction. The system was tested on the SROIE dataset and a private dataset to differentiate between genuine and fake billing documents. It was concluded that machine learning could detect manipulations in receipts and billing records. However, the approach is specialized towards retail bills and does not provide a general document verification framework capable of handling certificates, bills, and other official documents.

Benhamza, Djeffal, and Cheddad [5] investigated textual feature extraction and machine learning classification for scanned administrative document forgery detection. Their research focused on the fact that document manipulation can create statistical differences in text regions and scanning characteristics. The proposed approach extracted textual intensity-related features and used machine learning classification for detecting forged administrative documents. Their datasets were scanned administrative documents, proving that analysis of such documents' text and structure is more informative than just the visual assessment. The study also demonstrated that efficient document-level analysis is possible without a reference original document for every document. However, the approach is focused towards a particular type of statistical textual analysis and does not incorporate OCR, multiple forgery categories, signature analysis, and cryptographic fingerprinting in one system.

More recent research has been focusing specifically on academic certificate verification using hybrid deep learning. In 2026, IEEE research proposed a hybrid model based on ResNet50, EfficientNetB0, Vision Transformers, convolutional autoencoders, and Siamese networks for detecting academic certificate forgeries. The system considered pixel-level

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

inconsistencies, global document context, new types of manipulation patterns, and similarity of signatures and stamps. This shows the trend towards multiple machine learning models for improving certificate authentication but at the expense of increased computational and implementation complexity, which may not be appropriate for light academic projects or systems meant for straightforward document verification.

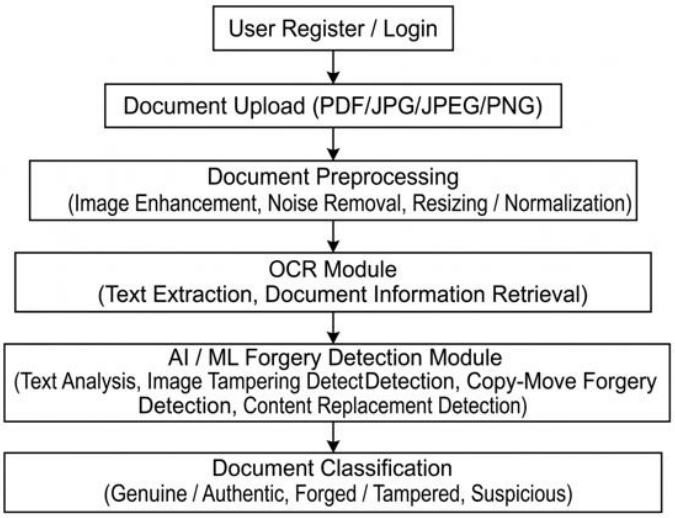
The reviewed articles confirm that single methods could solve specific tasks related to detecting document forgeries. For example, CNN- and deep learning-based techniques proved their efficiency in detecting of visual and structure manipulations, while the methods using OCR demonstrated the ability to extract and analyse text data elements. Certificate-specific algorithms were mainly used to solve academic document manipulations, while special software was developed for bills and receipts. Cryptographic hashing provides a different type of protection, detecting changes to the digital content of a document. These approaches demonstrate that document authenticity cannot be determined effectively with a single technique.

However, there is an important research gap related to integrating these complementary techniques into one document verification system. Many existing studies focus on one category of documents, one type of forgery, or one detection approach. For example, some systems emphasize image-level manipulation, whereas others focus on textual characteristics, certificates, or bills. Hash-based methods could detect changes, but they do not identify the nature of the forgery, while machine learning methods could detect suspicious patterns but do not necessarily provide a persistent digital fingerprint for subsequent integrity comparison.

The proposed AI Powered Document Forgery Detection and Intelligent Verification System Using Machine Learning addresses this research gap by combining multiple stages in a workflow. The system accepts certificates, bills, and other official documents in PDF and common image formats. Receipts and bills were initially subject to preprocessing followed by the application of OCR to extract necessary data. Furthermore, computer vision and machine learning techniques detected textual manipulations, image forgeries, copy-move, irregular fonts, and signature counterfeits. Finally, the system classified the document as Genuine/Authentic, Forged/Tampered, or Suspicious with a specific probability. Besides, the study also involved advanced techniques, such as creating a SHA-256 digital fingerprint stored as a hash along with the document data and prediction results. Thus, the system could track document changes by comparing two fingerprints and analysing the dissimilarities. Therefore, the proposed research extended the scope of AI-driven document forgery detection methods by using cryptographic hash functions to store vital information about the document's authenticity and track its changes.

3. METHODOLOGY:

SYSTEM WORKFLOW



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

3.1 User Authentication

The system will first present a login screen. Authentication will make sure that the document's verification functionality will be available to the user via the application's interface.

3.2 Document Upload

The user will then be able to upload the relevant digital document/s that is/are compatible with the application. The types of digital documents supported by the proposed system include PDF, JPG, JPEG and PNG. These documents will form the input of the verification pipeline.

3.3 Document Pre-processing

Preprocessing is an important step as the quality of the input document can influence the performance of following steps like OCR and image analysis.
The document is converted into a suitable form for further processing. Image processing such as sharpening and noise reduction can be performed at the desired places. The aim is to obtain a document with clear content for the following procedures.

3.4 Optical Character Recognition

OCR will extract any textual information contained in the image of the document. The extracted text provides machine-readable information that can be utilized during document analysis.
OCR also allows for the system to incorporate the information extracted from the document into the verification report.

3.5 Forgery Feature Analysis

The proposed system will take into account multiple features related to forging.
Text Forgery
Image Forgery
Copy-Move Forgery
Irregular Font Detection
Signature Forgery

3.6 MACHINE LEARNING AND COMPUTER VISION

Machine Learning will play a crucial role in the proposed system. The goal is to make use of document features to decide if a document is authentic or has been forged.
The system can utilize the document features extracted as described above as input into a classification model. The output of the model will provide a classification result and a corresponding score for the document. Therefore, using both technologies in combination will allow us to obtain a more accurate result than using text-based checks alone.

3.7 Document Classification

This is the **final decision stage** of the detection process.

Based on the results obtained from the AI/ML forgery detection module, the system places the document into one of three categories:

- Genuine / Authentic – the document does not show significant signs of forgery.
- Forged / Tampered – the system detects strong evidence of modification or manipulation.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

- Suspicious – the system finds some unusual features, but the confidence is not sufficient to clearly classify it as genuine or forged.

The system can also provide a prediction/confidence score to indicate how strongly the model supports its classification.

Evaluation metrics:

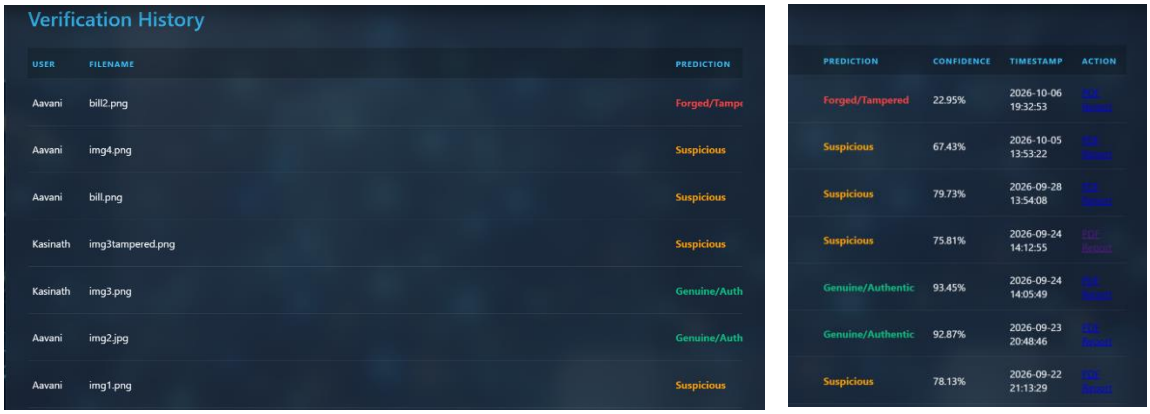
The proposed system is evaluated on the basis of Accuracy, Precision, Recall, F1-Score and Confusion Matrix to measure the effectiveness of the system in classifying the documents as Genuine, Forged or Suspicious.

Metric	Purpose	Formula
Accuracy	Measures the overall percentage of documents classified correctly.	(TP + TN) / (TP + TN + FP + FN)
Precision	Measures how many documents predicted as forged are actually forged.	TP / (TP + FP)
Recall	Measures how many actual forged documents are correctly detected.	TP / (TP + FN)
F1-Score	Provides a balanced measure of Precision and Recall.	2 × (Precision × Recall) / (Precision + Recall)
Confusion Matrix	Shows correct and incorrect classifications for Genuine, Forged, and Suspicious documents.	—

4. RESULTS AND DISCUSSION

The developed system was tested using different document files like bills and other document images. The verification history shows that the system is able to keep and show the results of the documents that have been analysed before. The verification records presented indicated that the system classified the test documents as Genuine and Suspicious, and the detailed verification page further indicated the detection of a Forged/Tampered document.

FIGURE 1: Verification history



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

Classification	Number of Documents	Percentage
Genuine	2	42.86%
Suspicious	5	57.14%
Total	**7**	**100%**

Looking at the verification history, out of the seven records shown, two were marked as Genuine and five were marked as Suspicious. This shows that the system is able to distinguish documents based on the identified features and to save the verification outcomes to refer to them later.

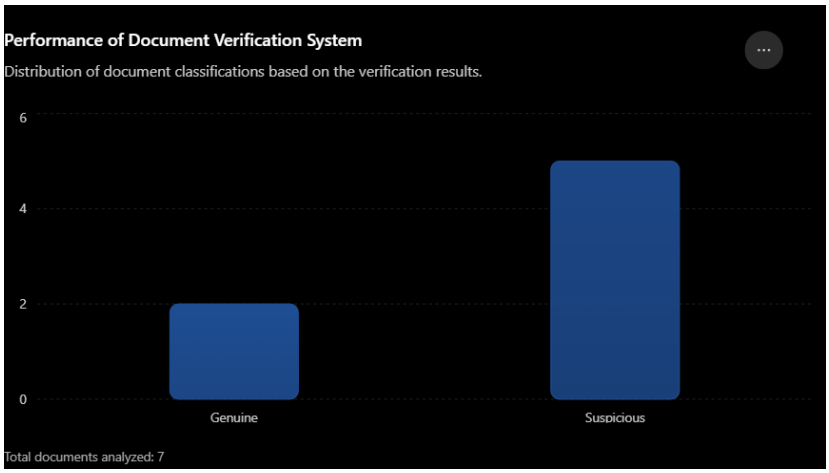
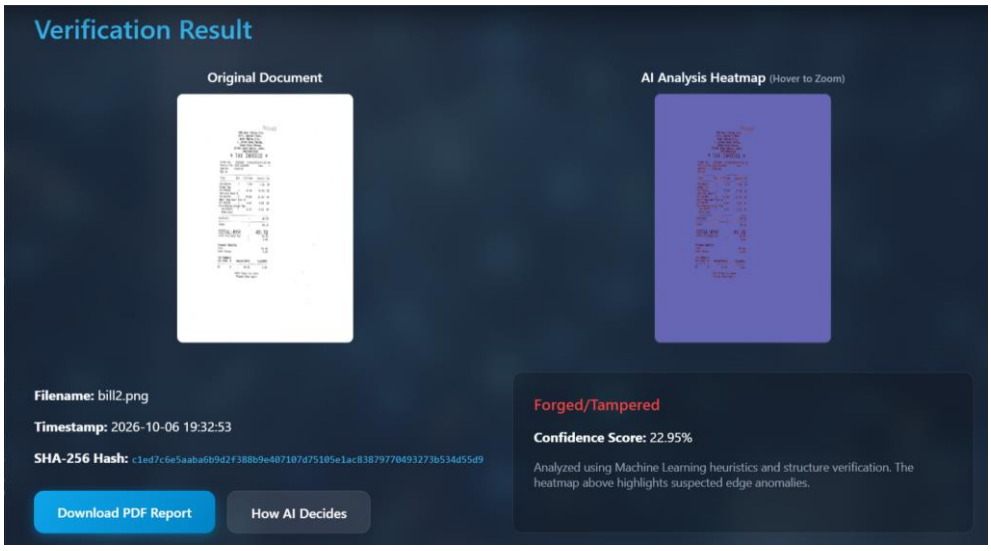


FIGURE 2: AI based verification result



The detailed verification result states bill2.png was identified as Forged/Tampered with a displayed confidence score of 22.95%. The system also generates an AI analysis heatmap to highlight the suspected document anomalies. In addition, a SHA-256 hash is generated for the document, offering a digital fingerprint that can be used to verify document integrity.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

5. CONCLUSION

Proposed AI Powered Document Forgery Detection and Intelligent Verification System Using Machine Learning Research Paper are a concept note aimed at solving the problem of identifying and verifying digital documents such as certificates, bills, and other official files. The system utilizes image preprocessing, OCR, machine learning, computer vision, and SHA-256 hashing to scan, analyse, and verify each document's contents, detect possible manipulations, and check the legitimacy of the document. The created system successfully classifies the documents into Genuine, Forged/Tampered, and Suspicious with respective prediction and verification results. The conclusion drawn from the research indicates that the proposed system would help speed up the verification process compared to the manual verification process currently in use. Therefore, the combination of the AI-based forgery detection system and cryptographic hashing will provide an extra level of document authentication and integrity checks.

6. FUTURE WORK

The future work on the proposed AI Powered Document Forgery Detection and Intelligent Verification System Using Machine Learning can be carried out in order to increase the accuracy, efficiency, and practicability of the current model and system. The upcoming future work can be performed in the following manner:

- Using a large range and variety of the datasets of the certificates, bills, receipts, and other official documents that can include both genuine and forged samples of the documents.
- Test other machine learning and deep learning algorithms to check which one can offer more accurate forging detection.
- Increase the prediction accuracy using advanced image pre-processing and feature extraction methods.
- Create a mobile application for the user to verify the documents conveniently.
- Implement real-time document verification for a faster detection and authentication process.
- Add more official documents and bills to verify and detect fake ones from different organizations.
- Develop the system using the concept of Explainable AI (XAI) to make the model able to explain or justify the authentication outcome.
- Improve the existing verification technique based on the SHA-256 algorithm to offer a more secure and scalable document integrity management.

7.REFERENCE

[1] M. Sirajudeen et al., "Forgery document detection in information management system using cognitive techniques," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 6, pp. 8057–8068, 2020.
View Article

[2] Y. Sun, R. Ni, and Y. Zhao, "MFAN: Multi-Level Features Attention Network for Fake Certificate Image Detection," *Entropy*, vol. 24, no. 1, p. 118, 2022.
View Article

[3] S. Boonkrong, "Design of an academic document forgery detection system," *International Journal of Information Technology*, vol. 17, pp. 5175–5187, 2025.
View Article

[4] D. Chakraborty, "Bill safe: intelligent forgery detection with CNN and upgrade sand cat swarm optimization," *Discover Computing*, vol. 28, article 84, 2025.
View Article

[5] H. Benhamza, A. Djeffal, and A. Cheddad, "Textual Feature Extraction and Machine Learning Classification for Document Forgery Detection," *Applied Computational Intelligence and Soft Computing*, vol. 2026, article 4463640, 2026.
View Article

[6] "Forgery Detection in Academic Certificates Using Hybrid Deep Learning Framework," *2026 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, 2026.
View Article

(This work is licensed under a Creative Commons Attribution 4.0 International License.)