Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 19, No. 1, 2025

Research on the Fusion of Data Analysis Methods and Artificial Intelligence in Electronic Information Systems

Jinzeng Zhu

School of Communication and Information Engineering, Chongqing College of Mobile Communication, Chongqing, China

Abstract: This research initially reveals that as digital transformation deepens across industries, electronic information systems have become core business hubs. The exponential growth of the global datasphere (IDC data: 120ZB in 2024, projected to exceed 150ZB in 2025) inherently places higher demands on the intelligence of the "data-decision-making" chain. The shortcomings of traditional statistical modeling and manual feature extraction are not only technical limitations but also hinder the industrys transition from "experience-driven" to "data intelligence-driven" decision-making. The integration of AI and electronic information systems is the key to overcoming this dilemma and reshaping decision-making efficiency. This article focuses on the theme of integration: first, analyzing the logical foundation and internal mechanisms to clarify the dimensions of AI empowerment; then, building a technical architecture and exploring data processing and algorithm adaptation paths; combining industrial fault diagnosis and financial risk assessment to verify effectiveness, simultaneously proposing data compliance and model risk management solutions, and ultimately refining optimization strategies. This study does not cover integration solutions for small and medium-sized enterprises with low computing power, and the applicability of the conclusions to hardware resource constraints remains to be verified. In the future, we can focus on the development of lightweight AI models (such as the optimized version of TensorFlow Lite) or launch low-cost integration toolkits to lower the threshold for intelligent transformation for small and medium-sized enterprises. This article finds that the integration of the two requires data quality as the cornerstone, algorithm innovation as the core, and compliance control as the barrier. It can not only improve the accuracy and real-time performance of system data analysis, but also promote its transformation from "data storage end" to "intelligent decision-making center", providing solid technical support for industrial innovation.

Keywords: Electronic information systems; Data analysis methods; Artificial intelligence; Fusion technology; Risk management; Data compliance.

1. Introduction

This research initially reveals that the demand for millisecond-level response times in financial anti-fraud scenarios exposes the limitations of traditional manual analysis. This stems from the inability of traditional data processing paradigms to adapt to the time-series data from industrial sensors and unstructured public administration data generated by 5G and the Internet of Things. The core driving force behind the transformation of electronic information systems from "business record-keeping tools" to "core data production" is the massive expansion of this type of data. The lag in traditional statistical analysis is a key bottleneck hindering the unlocking of datas value. AI, posing a breakthrough, boasts core strengths such as autonomous feature extraction. However, its integration with electronic information systems is more than a mere technological overlay; it is a core proposition that determines the depth of intelligent implementation across multiple industries. Existing research, however, suffers from significant gaps: Liu Haijuns macro -analysis fails to address specific data analysis scenarios, Xia Yuhangs fault diagnosis research fail s to cover the entire integration process, and Anuars compliance research fails to integrate with the analytical process. Most research is limited to a single technology or scenario, neglecting the full synergy of the "data-algorithm-application-compliance" chain and weakening the fundamentals of data processing. The value of electronic document management, emphasized by Sprague Jr. R.H. in 1995, remains insufficiently recognized in this data-intensive era. This directly leads to the frequent dilemma of "the technology works, but the processes dont flow" when implementing integration solutions. This study attempts to focus on the data analysis of electronic information systems, integrate the results of federated learning and hybrid algorithms, and construct a research framework (including logical mechanisms, technical architecture, etc.) from five dimensions to provide theoretical and practical support for intelligent upgrades. However, this study does not consider the impact of extreme high-temperature and high-dust industrial conditions on data collection quality, and the applicability of the conclusions to such scenarios remains to be verified. In the future, we can focus on developing data preprocessing solutions for extreme working conditions, and at the same time work with governments and enterprises to build cross-industry "data-algorithm-compliance" collaborative standards to further lower the threshold for integration and implementation.

2. The Logical Basis for the Fusion of Data Analysis and AI in Electronic Information Systems

This research has initially revealed that industrial edge gateways are reshaping the data processing paradigm. By deploying lightweight deep learning models (such as TensorFlow Lite-optimized models), they enable real-time analysis of device data and local decision-making, breaking the latency bottleneck of traditional remote architectures. The value of this breakthrough was demonstrated by the fact that a certain automotive companys welding workshop experienced missed fault notifications due to this latency. This restructuring also signals a crucial shift in industrial data processing from "cloud-dependence" to "edge autonomy," providing core support for high-real-time operations and maintenance. Embedding AI alone doesnt solv e all problems. The essence of integrating AI with edge gateways lies in the adaptation of business and technology. Low power consumption requirements at the edge dictate that models cannot have too many parameters. Traditional regression analysis suffers from the "curse of dimensionality," further highlighting the need for adaptation. The applicability of these findings beyond multi-node scenarios remains to be verified, as this research does not address the issue of computing power allocation for multi-gateway collaboration. Future research could focus on optimizing multi-gateway collaboration through the integration of federated learning and developing low-power models compatible with ARM chips. Static analysis of time-series data from financial transactions cannot capture dynamic trends, significantly reducing predictive accuracy. Furthermore, in fault diagnosis scenarios, engineers rely on empirical analysis to label fault characteristics, which is not only inefficient but also susceptible to subjective judgment. These pain points just provide a space for AI to develop - machine learning can automatically filter key information from high-dimensional data, deep learning models such as LSTM and Transformer are good at grasping time series patterns, and reinforcement learning can also optimize models through autonomous iteration, greatly reducing manual intervention. However, the release of AIs value is inseparable from the support of electronic information systems. Generally speaking, AI models require massive amounts of high-quality data as "fuel", and the systems database happens to store information on the entire business process; and the systems industrial control, financial transactions and other scenarios provide AI with a clear goal orientation, preventing it from falling into "technical idleness without scenarios". When Liu Haijun studied the integration of AI and industrial innovation, he pointed out that "the deep binding of technology and scenarios is the core mechanism of integration, and AI separated from scenarios is difficult to generate actual value" [1]. This view also holds true in the data analysis link - only by combining AI algorithms with specific needs such as system fault warning and risk identification can the technology be truly implemented. To a certain extent, this integration is also an inevitable result of technological evolution: as the complexity of electronic information systems increases, data analysis needs have shifted from "post-statistics" to "pre-prediction" and "real-time decision-making", and lightweight deployment of AI, edge AI and other technological iterations just meet this demand. It can be said that it is the needs of the system that push AI forward, and it is the capabilities of AI that force the system to upgrade. The closed loop formed by the two is the true core of integration.

3. Core technical architecture and data processing methods of integration When a cross-regional

Power grid built an AI load forecasting system, it was once stagnant because the data in each region was not shared with each other. It was not until the introduction of federated learning that different nodes could jointly train the model without transmitting the original power consumption data, which not only protected privacy but also improved accuracy. This just confirms Sprague Jr R Hs judgment: the value of the information system comes from data quality, and in the AI integration scenario, this dependence is even greater. After all, AI is much more sensitive to poor-quality data than traditional analysis methods [2]. But can the effective implementation of integration be ensured by simply dividing it into "data layer-algorithm layer-application layer"? I am afraid not. The author believes that the core of the architecture is the adaptation of each link scenario, rather than mechanical stratification. Taking data processing as an example, unstructured documents that are easily overlooked in electronic information systems, such as equipment maintenance records, contain key information. They must first be converted into electronic text using optical character recognition, and then extracted by natural language processing to extract the fault time and location. Finally, they must be standardized and stored before they can be linked with other data. Generally speaking, relational databases are a safe choice for structured data, while distributed file systems are suitable for unstructured documents or audio. Time series data requires a time series database. The choice of algorithm should also be closely aligned with data quality and the application scenario: for equipment fault classification, random forests or support vector machines are sufficient; for time series predictions like power grid load, LSTMs or GRUs are more optimal; for complex problems like financial credit assessment, an XGBoost-MLP hybrid model is more suitable. For industrial edge deployment, it is advisable to choose lightweight models such as MobileNet and adapt to hardware support; the application layer must adhere to the "human-machine collaboration" principle. For example, when converting fault diagnosis results into maintenance instructions, human decision-making is essential, which can, to some extent, mitigate the risks of purely AI-based decision-making. Ultimately, a technical architecture is not a static framework but a dynamically adaptable system. Data processing, from noise reduction and completion to dimensionality reduction, paves the way for the algorithm layer. The integration of algorithms and applications must be based on the business scenario. An architectural design that is divorced from practical needs, no matter how sophisticated, is merely theoretical.

4. Practice of fusion algorithm based on deep learning The rotating machinery fault diagnosis

Experiment of Xia Yuhang et al. is very inspiring: the diagnostic accuracy of its convolution residual LSTM model exceeds 98.5%, which is much higher than the 85% of traditional vibration signal spectrum analysis. Rotating machinery is the core equipment of industrial electronic information system. Fault diagnosis has long relied on manual extraction of vibration signal features and expert experience judgment, which is inefficient and subjective. The "feature extraction-fault classification" end-to-end analysis of this model just hits the pain point of traditional methods [3]. But does this mean that traditional analysis methods have lost their value? Not necessarily. It is just that when processing highly complex data, its shortcoming of relying on manual experience is more prominent. The author believes that the core advantage of this model is to jump out of single feature extraction and realize the coordinated capture of spatial and temporal features: using CNN convolution layer to process the time domain or spectrum of the vibration signal to lock the abnormal area of the fault waveform; using residual connection to solve the CNN gradient disappearance and ensure feature transmission; then inputting spatial features into LSTM, using its gating mechanism to capture signal temporal changes, and relying on shared weights to reduce parameters to improve efficiency. This research has initially revealed that deploying models tailored to industrial scenarios within electronic information systems is a key enabler for the implementation of industrial intelligence, and goes beyond simply grafting on new technologies. Edge node computing power requires hardware support (such as an Intel Core i7 processor and 8GB+ of memory). Embedding fault diagnosis modules using lightweight models such as TensorFlow Lite meets the timeliness requirements of diagnosis. This research did not consider the impact of high-temperature operating conditions on hardware computing stability. Future efforts could focus on optimizing hardware solutions by integrating low-power embedded chips to reduce energy consumption. At a 10dB signal-to-noise ratio, the accuracy was 96.2%, surpassing the 89.5% of a traditional CNN and the 87.3% of a single LSTM. Single-group signal analysis takes only 20ms, meeting the sub-100ms real-time requirements of industrial systems. This adaptability can be extended to applications in transportation (congestion prediction) and healthcare (assisted diagnosis of heart disease). The author believes that these scenarios all require the simultaneous processing of spatiotemporal data, and the algorithm is implemented precisely to meet this need. Even the most sophisticated optimizations that are not considered relevant to the scenario are of little practical value.

5. Risk Assessment and Model Optimization in Fusion Applications

The credit assessment module of a commercial banks digital supply chain finance system once encountered a typical technical adaptation dilemma. When analyzing corporate transaction flows and financial statement data, the models under-identification rate for high-risk customers climbed to 18% due to over 85% of samples being "high-quality customers" and less than 15% being "risky customers." Despite subsequent replenishment of risky samples through an industry data alliance, the model suffered from overfitting due to overlearning the training set features. While the training set accuracy reached 95%, the accuracy of assessments for newly added customers plummeted to 75%. This misjudgment led to the exposure of three delinquent loans in a batch of small and micro-enterprise loans. This study argues that this case highlights the triple risk transmission chain of the integration of electronic information systems and AI: sample distribution biases at the data level directly affect the generalization capabilities of the algorithm layer, ultimately transmitting to the application layer and weakening the reliability of business decisions. This research has initially revealed that simply adjusting sample size or model parameters cannot completely mitigate risks in industrial scenarios. The XGBoost-MLP hybrid model proposed by Li Y et al. (2023) is a key path to balancing data processing accuracy and business adaptability. This study did not consider the impact of extreme temperature and humidity on model stability. Future research could focus on optimizing the models real-time responsiveness by integrating it with edge computing [4]. The core logic of this model lies in "step-by-step decomposition of risk identification tasks": first, using the feature importance ranking function of XGBoost, six key risk factors such as the companys debt-to-asset ratio and cash flow turnover rate are screened out to reduce data bias caused by redundant features; then, using the nonlinear fitting ability of MLP, the correlation between transaction frequency fluctuations and implicit credit risk is captured. From the perspective of performance indicators, the core indicator of the risk assessment model, AUC, reached 0.92, higher than the 0.86 of XGBoost and the 0.88 of MLP; more importantly, 6 months of business tracking showed that its performance decay rate was only 5%, far lower than the decay range of 8%-10% of the single model, and its long-term stability was significantly improved. However, model optimization is not a one-time solution. This study found that a dynamic control mechanism is also needed: regularly update the training set with new data accumulated in the electronic information system to avoid model drift; form a joint technical-business team to audit the results, and adjust the parameters in a timely manner when the accuracy rate falls below the 85% threshold; and set up a manual review process for high-risk businesses. After all, the risk of integrating the two is essentially a balance between technical adaptability and business security. We believe that technological optimization and mechanism guarantee must be approached simultaneously.

6. Data Compliance and Audit Mechanisms in Integration Scenarios

This study noted that a medical electronic information system had directly shared patient medical records with a third-party AI company for model training. This system failed to desensitize sensitive information like names and ID numbers, and failed to inform patients of the intended use of the data during the collection phase. This practice violates the "informed consent" principle of the Personal Information Protection Law and infringes upon GDPRs data minimization and privacy protection requirements. It also exposes the common pain point of the integration of electronic information systems and AI: in the flow of massive data, compliance often becomes an “accessory” to the advancement of technology. Anuar N B’s research has pointed out that while AI technology improves data processing efficiency, it actually amplifies compliance risks, and a dual control system needs to be established through technical means and institutional design [5]. But does this mean that as long as AI compliance tools are piled up, risks can be completely avoided? Obviously not. Although a European financial institution uses AI-driven compliance detection tools to increase data compliance detection efficiency by 60% and reduce the false positive rate to less than 3% - such tools use natural language processing to analyze collection protocols and monitor data flows, and can promptly identify problems such as “exceeding the scope of collection” and “inconsistency between use and declaration” - technology ultimately has its limits. For example, it cannot solve institutional problems such as “ambiguous attribution of compliance responsibilities”. In terms of data desensitization and anonymization, differential privacy technology adds subtle noise to data, preventing third parties from inferring user identities. This research has initially demonstrated the value of federated learning—it achieves "data availability without visibility" through local training and parameter-only transmission. It not only preserves data value and reduces the risk of privacy leaks, but also overcomes the traditional conflict between data sharing and privacy protection, providing a viable path for cross-institutional data collaboration in sensitive industries such as finance and healthcare. Compliance management for the integration of electronic information systems and AI must be centered around "technical support + institutional constraints"—a single technology alone cannot provide a complete defense. Cases where some companies neglect institutional development, leading to compliance failures, reinforce the necessity of synergy between the two. This study has limitations: it does not consider the impact of heterogeneous data such as medical images and industrial sensors on compliance process adaptation, nor does it cover low-cost compliance solutions for small and micro enterprises. Future research focuses on two areas: optimizing the security of federated learning parameter transmission through the integration of differential privacy, and developing a lightweight compliance assessment module for the manufacturing industry to lower the compliance threshold for small and medium-sized enterprises.

7. Conclusion

The research by Xia Yuhang and others not only demonstrates the advantages of the convolutional residual LSTM model in a certain heavy industry operation and maintenance scenario - the diagnosis accuracy of bearing wear and gear tooth breakage is 98.5%, significantly better than the 82% of traditional spectrum analysis, but also preliminarily confirms that the integration of electronic information systems and AI needs to be centered on "demand complementarity and capability matching": machine learning makes up for the shortcomings of traditional analysis and processing of high-dimensional time series data, and the system provides a reliable data carrier for AI, providing a practical path for industrial operation and maintenance decision optimization. This study did not include the interference of high dust conditions on the model input data, and the applicability of the conclusions in this scenario remains to be verified. In the future, we can focus on combining edge gateways (such as the EG-300 model) for lightweight model deployment and expand to the diagnosis of multiple types of rotating machinery such as fans and motors. Liu Haijuns pr oposed "Logic of Integration of AI and Industrial Innovation" is here visualized as a collaborative closed loop of "data-algorithm-scenario." However, this closed loop is not applicable to all systems. The implementation of this integration requires data processing as a prerequisite: Sprague Jr. R. H.s 1995 emphasis on electronic document management has now expanded to encompass "high-quality processing of all types of data." From standardized sensor data collection to federated learning collaborative training, each step impacts AI performance. Algorithm scenario adaptation and application-level "human-machine collaboration" are also key to mitigating the risks of pure AI decision-making. Risk management and compliance are often overlooked but crucial. Referring to Anuar N Bs empirical research on AI compliance, technical risks such as data bias and model drift can be mitigated through hybrid algorithms, while compliance risks require AI data desensitization tools and a "full-process model accountability system + third-party evaluation" approach to manage and control. This study has limitations: it does not address lightweight model deployment at the edge or the integration costs for small and medium-sized enterprises (SMEs), who have limited hardware and are far more cost-sensitive than large enterprises. In the future, large models may drive integration towards "scenario-based general intelligence," with cross-domain collaboration at the core and the joint efforts of government, enterprises, and academia to build an ecosystem to promote system intelligence.

References

[1] Liu Haijun. Artificial intelligence drives the integration of scientific and technological innovation and industrial innovation: logic, mechanism and path [J]. China Circulation Economy, 2025, 39(09): 3-14. DOI: 10.14089/j.cnki.cn11-3664/f.2025.09.001.

[2] Sprague Jr R H. Electronic document management: challenges and opportunities facing information system managers [J]. MIS Quarterly, 1995: 29-49.

[3] Xia Yuhang, Chen Xing, Fu Wenlong, et al. Rotating machinery fault diagnosis based on convolutional residual shared weight LSTM [J]. Machine Tools & Hydraulics, 2023, 51(16): 215-221.

[4] Li Y, Stasinakis C, Yeo W M. Hybrid XGBoost-MLP model for credit risk assessment of digital supply chain finance [J]. Forecasting, 2022, 4(1): 184-207.

[5] Anuar N B. The role of artificial intelligence in GDPR compliance and data protection audit [J]. Multidisciplinary Innovation and Research Analysis, 2023, 4(4): 1-15.