K-Nearest Neighbor-Based Detection and Mitigation of False Data Injection Attacks in Nonlinear Automatic Generation Control Systems

Xin Ge
Information Office, University of Shanghai for Science and Technology, Shanghai 200093, China
E-mail: gexin_usst@163.com

Keywords: detection mechanism, K-nearest neighbors (KNN), automatic generation control (AGC), false data Injection (FDI), power system

Received: May 21, 2025

The security of Automatic Generation Control (AGC) systems is crucial for maintaining frequency stability in modern power grids. False Data Injection (FDI) attacks can compromise AGC operations by altering measurement data and misleading control actions. This paper proposes a K-Nearest Neighbor (KNN)-based detection and mitigation framework for FDI attacks, explicitly considering AGC nonlinearities such as the governor dead-band (GDB), generation rate constraint (GRC), and transportation time delay (TTD). The proposed non-parametric model detects abnormal data by analyzing feature distances without requiring extensive training data or computational resources. A two-area AGC test system is used for validation, and the method is evaluated in terms of detection accuracy, false positive rate (FPR), and computational efficiency. Simulation results demonstrate that the proposed approach achieves a detection accuracy of 95.2% and an FPR of 3.5% at k = 5, while reducing computation time by more than 80% compared to deep learning methods. These findings confirm that the KNN framework offers a lightweight and effective solution for real-time FDI attack detection and mitigation in nonlinear AGC systems.

Povzetek: Prispevek predstavi lahek KNN-pristop za zaznavanje in ublažitev FDI napadov v AGC sistemih, ki učinkovito izboljšuje varnost in stabilnost elektroenergetskih omrežij.

1 Introduction

AGC systems help power grids stay steady, making sure generator power matches up with what people need [1], [2]. They are, however, open doors for cyber threats like FDI attacks [3], [4]. In those, bad data tricks AGC to mess up the whole power grid control [5]. And why? Because handling these threats is a must; an FDI hit can throw things off a little or, in the worst case, mess up the grid big-time.

As more communication and tech stuff goes into power systems, AGC systems get even more open to cyber tricks [6]. These attacks dodge the usual detection methods, showing how important it is to make better ways to spot them [7]. With all that, this work is all about keeping power grids strong and steady without piling up on the computer work needed to stop these cyber tricks. The detection of FDI attacks in power systems, especially within the domain of the AGC, has attained considerable attention in the last decade. Researchers have tried a variety of methodologies for the same problem, ranging from conventional statistical techniques to advanced machine learning models, but the DL approaches have emerged as favorites due to their ease in handling complex, high-dimensional data [8]. Some of the DL architectures tried on this problem include CNNs and RNNs. For instance, people have used CNNs because they can process spatial data and detect anomalies by learning the spatial correlations between power system measurements [9]. We prefer RNNs because they can learn the temporal dependencies inherent in time-series data, making them suitable for AGC systems where time is a crucial factor [10]. However, most of these DL-based techniques are computationally expensive and typically require huge amounts of labeled data, which may be difficult to obtain in a power system.

Hybrid approaches, which combine traditional signal processing methods with machine learning techniques, have recently emerged in response to the challenges associated with pure DL models. The rationale behind these hybrid models is to exploit the robustness of statistical methods along with the pattern recognition capability of machine learning methods [11]. Some works have integrated the Kalman filter with neural networks to improve the accuracy of FDI attack detection while reducing the computational load [12]. Other concepts utilize Principal Component Analysis (PCA) for data preprocessing and employ Support Vector Machines (SVM) to enhance the detection algorithm [13]. While these methods offer a good balance between accuracy and efficiency, they also face challenges in real-time application due to the need to fine-tune multiple techniques across various systems [14].

But with these better detection methods available, the number of practical applications of these techniques to real-world AGC systems is still limited. In fact, one of the important barriers to deploying these techniques lies in the heterogeneity of power system configurations and

operational characteristics in different grids. The development of these models focuses on specific system configurations, making their extension to other setups challenging due to the high cost of retraining or recalibration [15]. The presence of inherent nonlinearities in an AGC system (which includes those like GDB, GRC, and TTD) adds to the complexity [16]. These nonlinearities mask the signatures of FDI attacks, thereby making the detection process more difficult [17]. In light of these factors, most current detection schemes lose their effectiveness, leading to an emerging need for more adaptive and robust detection techniques [18].

Despite significant advancements in detecting FDI attacks in AGC systems, certain research gaps persist. While most current detection approaches, particularly those based on DL, guarantee high detection accuracy, they typically come with high computational costs [19]. These methods usually need huge training datasets and high computational resources to work properly, making them less practical for real-time applications in the AGC systems that require quick and sure responses against probable threats. Furthermore, these models do not exhibit significant adaptability across various power system configurations [20]. Many DL-based models, trained on specific datasets and system characteristics, will encounter numerous challenges when applied to grid conditions with varying operational dynamics or configurations. Obviously, this constitutes one of the main barriers to their widespread deployment across varied AGC systems.

In addition, other critical challenges in FDI attack detection are intrinsic nonlinearities within the AGC systems, like GDB, GRC, and TTD. These nonlinearities complicate the attack detection process by introducing variability in system behavior, which could potentially mask or mimic the effects of the FDI attack. Most of the current detection frameworks fall short in this regard; there is a lot of non-linearity that is often hard to deal with, thus reducing the reliability and accuracy in real-world applications. This necessitates the development of detection techniques capable of handling complex tasks with minimal computational burden and maximum versatility within the system configuration.

This research addresses the identified gaps by introducing the K-nearest neighbors' algorithm as a novel approach to detecting FDI attacks in AGC systems. The main contribution of the present work is in developing and applying KNN, a simpler yet powerful machine learning technique that offers a number of advantages compared to more complex DL-based methods. Unlike deep learning models, KNN does not require intensive computation or large data sets to be trained; hence, it is very suitable for real-time applications in AGC systems. KNN, using a distance-based way, can mark data as normal or messed up without any fancy training or tuning. This makes it easy to use on all kinds of AGC setups and works in lots of situations. Here, this research puts KNN together with a model that has nonlinear AGC parts like GDB, GRC, and TTD, helping it find issues even with tricky AGC stuff that most other setups can't handle.

Unlike existing machine-learning-based FDI detection schemes that rely on linearized AGC models or ignore inherent nonlinear characteristics, the proposed KNN framework explicitly integrates governor dead-band (GDB), generation rate constraint (GRC), and transportation time delay (TTD) as descriptive features. This integration allows the detector to capture practical AGC dynamics under both normal and perturbed conditions. Consequently, the framework offers a distinct balance between computational efficiency, interpretability, and adaptability for real-time applications, providing a novel perspective on lightweight FDI detection compared with SVM-, PCA-, or deep-learning-based approaches.

This work is motivated by the following research questions:

(i) Can a K-Nearest Neighbor (KNN)-based approach effectively detect False Data Injection (FDI) attacks in nonlinear Automatic Generation Control (AGC) systems?

(ii) How does the proposed method compare with established machine learning methods such as Support Vector Machines (SVM) and Deep Learning (DL) approaches in terms of detection accuracy, false positive rate (FPR), and computational efficiency?

(iii) Can the KNN framework maintain robustness in the presence of AGC-specific nonlinearities such as governor dead-band (GDB), generation rate constraints (GRC), and transportation time delay (TTD)?

The test model is the two-area AGC system. Turns out, KNN catches problems fast and doesn't trip up on AGC's nonlinear parts, so it works better than heavy DL-based ways. The paper kicks off with a look at why AGC is key in power systems and the rising danger of FDI attacks. These attacks are explained with numbers and how they shake system stability. Section II dives deep into the FDI attack model on AGC. Section III tells all about KNN detection and fixing tricks, its setup, and how well it works live against attacks. In Section IV, we go over results, weighing how the KNN way stands up to SVM and deep learning, checking for catch rates, false alarms, and speed. Wrapping up in Section V, there's a recap of what we learned, plus ideas on keeping AGC safer and where detection tools should go next in power systems.

2 False data injection attack model on automatic generation control (AGC)

FDI attacks bring big risks, shaking up the AGC system's balance and making power and load matching harder. Here, we look deep into modeling AGC, picking apart key nonlinear parts, setting up math for the FDI attack, and checking what happens with AGC's tricky behaviors.

2.1 AGC nonlinearities and system model

Automatic Generation Control (AGC) design makes sure frequency stays stable, and tie-line power stays where it should, keeping up when loads or power-making changes. Some big nonlinear parts in AGC are the governor dead-band, generation rate limits, and the power delay time, which depends on valve spots. For AGC, these need to be modeled well to understand how it acts, especially if

cyber-attacks show up. GDB represents a non-linearity that builds up a dead-band range of frequency deviations inside which the governor does not move. The idea behind this mechanism is to avoid unnecessary movements of the governor for minor variations in frequency that may cause wear and tear in its mechanical elements. You can represent GDB mathematically as follows:

$$\Delta f_{db}(t) = \begin{cases} 0 & if\ |\Delta f(t)| < \Delta f_{db} \\ \Delta f(t) & if\ |\Delta f(t)| \geq \Delta f_{db} \end{cases} \quad (1)$$

To prevent the governor from responding to minor and inconsequential frequency deviations, set it at a very meager magnitude. Figure 1 depicts that dead band delays effective action against frequency changes by introducing a zone of insensitivity in the system response.

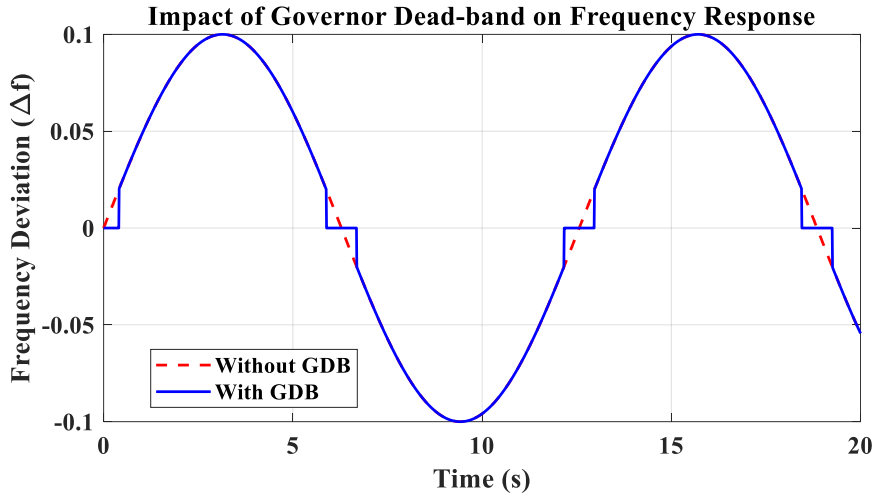

Figure 1: Impact of governor dead-band on frequency response

GRC limits the rate at which a generator's output can change. It is one of the key constraints, since much larger and quick changes in the generator output can cause excessive wear and tear on mechanical components of generators. Conventionally, we impose the GRC as the maximum rate of change, $R_{\max}$, on the generator output.

$$\frac{dP_g(t)}{dt} = \begin{cases} R_{\max} & \text{if } \frac{dP_g(t)}{dt} > R_{\max} \\ \frac{dP_a(t)}{dt} & \text{if } -R_{\max} \leq \frac{dP_g(t)}{dt} \leq R_{\max} \\ -R_{\max} & \text{if } \frac{dP_g(t)}{dt} < -R_{\max} \end{cases} \quad (2)$$

Here, P_g is the generator power output, and R_{max} is the maximum allowable ramp rate. The GRC thereby constrains the generated power output variations within a specified value and prevents sudden, potentially destabilizing variation in generation.

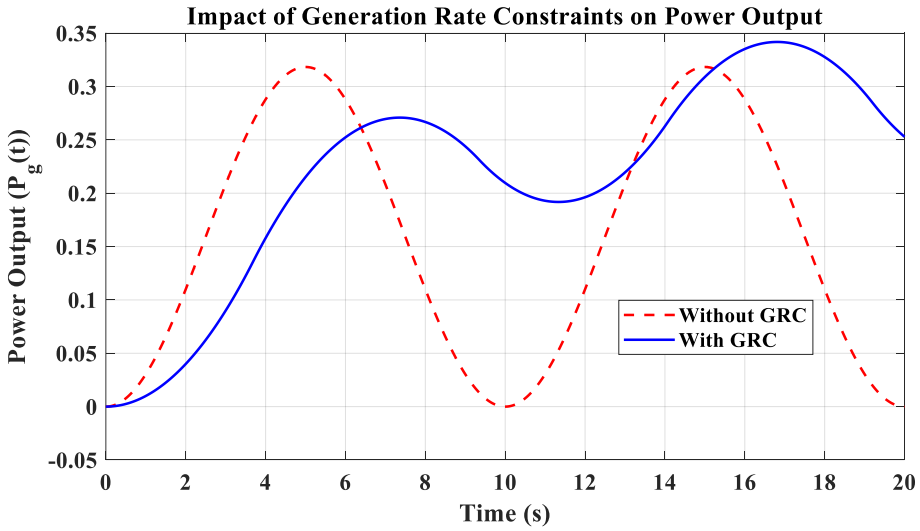

Figure 2: Impact of generation rate constraints on power output

Figure 2 illustrates the power output of a generator in response to a potential FDI attack. In other words, GRCs that limit the rate of system response lengthen the duration of the system's compromised state. Transportation time delay: There is an inherent time delay in communication, especially between the different components of the AGC system (basically, between the control centers and generators). Certainly, the transmission and processing of

data takes some time and affects the timing and effectiveness of the control actions. We can model this mathematically for TTD:

$$\Delta P_g(t) = \Delta P_c(t - \tau) \quad (3)$$

where τ is the delay time; $\Delta P_c(t)$ is the control signal from the AGC system. In this manner, the TTD will introduce a delay in detecting a frequency deviation or tie-line power flow error, thereby enabling the appropriate control action. Figure 3 depicts the effect of the TTD in the control signal and system response, especially during an FDI attack.

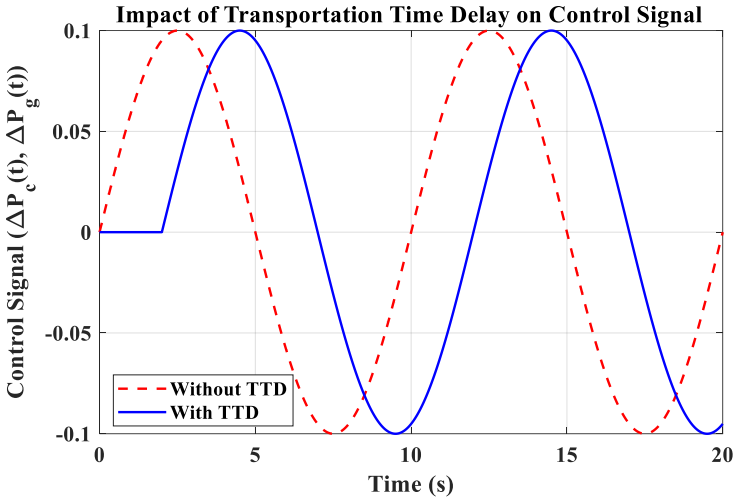

Figure 3: Impact of transportation time delay on control signal

This will illustrate how TTD can potentially affect both the control signal and system frequency in case of an FDI attack. This increases system vulnerability because the attacker has more time to execute their plan before control actions kick in. Nonlinear stuff makes it tougher for AGC to react right during an attack. Figuring out these nonlinear things is key to making models strong enough to see how systems act, both normally and when they're under attack.

2.2 Attack model

FDI attacks mess with AGC's control by putting wrong data into the measurements it depends on. When these measurements get messed up, the AGC might send out wrong commands, which could shake up the whole power grid. The attacker can play around with main AGC measurements like frequency $\Delta f(t)$ and tie-line flow $\Delta P_t(t)$ to cause trouble. In this setup, AGC shows how the false data messes up the readings like this:

$$\begin{aligned} \tilde{\Delta f}(t) &= \Delta f(t) + a_f(t) \\ \tilde{\Delta P}_t(t) &= \Delta P_t(t) + a_t(t) \end{aligned} \quad (4)$$

In this case, $a_f(t)$ and $a_t(t)$ are the extra bad data slipped into the frequency and tie-line power flow measurements. These fake numbers lead the AGC to make a messed-up control signal, called $\Delta\tilde{P}_c(t)$, calculated like this:

$$\tilde{\Delta P}_c(t) = K_p\left(\tilde{\Delta P}_t(t) + \frac{1}{s}\tilde{\Delta f}(t)\right) \quad (5)$$

Substituting the false data into this control law gives:

$$\begin{aligned} \tilde{\Delta P}_c(t) = K_p\Bigg(&\Delta P_t(t) + a_t(t) \\ &+ \frac{1}{s}\left(\Delta f(t) + a_f(t)\right)\Bigg) \end{aligned} \quad (6)$$

The formula here shows how those sneaky attack signals, $a_f(t)$ and $a_t(t)$, mess up the AGC's actions. With these wrong signals, the system could end up with big jumps in frequency and tie-line power, which may throw off the whole grid. When the attack hits, AGC either over or under its task, and that makes errors stick around or even start a back-and-forth swing in the system.

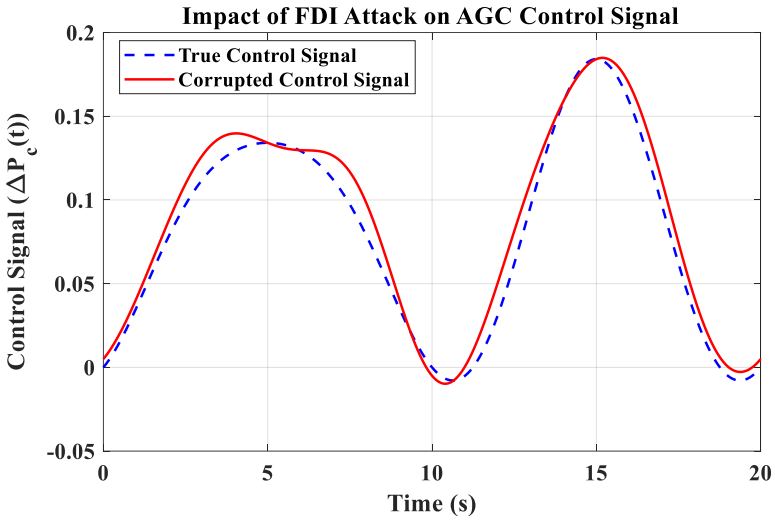

Figure 4: FDI Attack Model on AGC System

Figure 4 shows a simplified idea of an FDI attack. It shows where the attacker inserts data and how those fake measurements move through the AGC control loop and change the stability of the system.

1.1 AGC nonlinearities during cyber attacks

The AGC system incorporates nonlinearities such as the governor dead band, generation rate limits, and transportation time delay. These nonlinearities significantly impact the system's response to an FDI attack. Such nonlinearities serve to either magnify the effects of an attack or mask its detection, making the system more susceptible to sustained disruptions. The GDB might make it take longer for the AGC system to respond to an FDI attack because it doesn't pay attention to small changes in frequency that happen in the dead-band range. This causes the attack to last longer before AGC even starts to react. So, when the fake data eventually pushes frequency past the dead-band limit, AGC might respond too late, or it just doesn't fix things right. Here, the governor's reaction to an FDI attack with GDB in play can be shown like this:

$$\Delta f_{db}^{\tilde{}}(t) = \begin{cases} 0 & if\ |\tilde{\Delta f}(t)| < \Delta f_{db} \\ \tilde{\Delta f}(t) & if\ |\tilde{\Delta f}(t)| \geq \Delta f_{db} \end{cases} \quad (7)$$

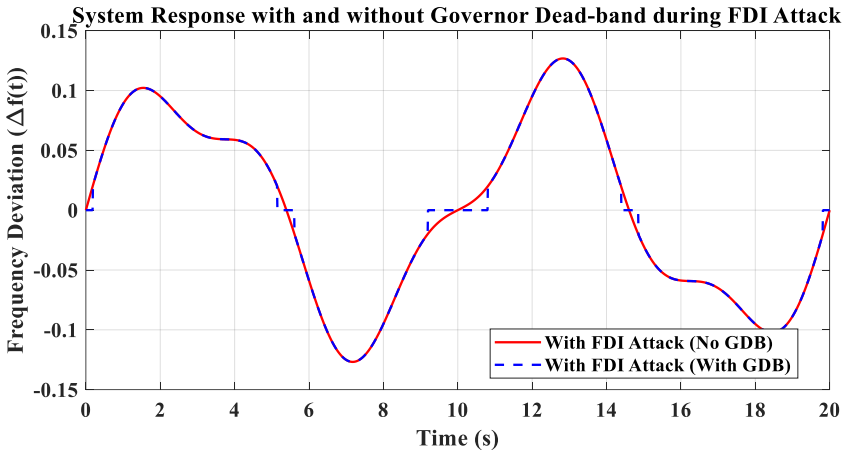

Figure 5: System response with governor dead-band during FDI attack

Figure 5 shows what happens to the AGC frequency response when there's an FDI attack, comparing it with and without GDB. It shows how an attacker might use the dead band to keep the attack going.

3 Attacks detection and mitigation using K-Nearest Neighbors (KNN)

The innovation of the proposed method lies in adapting the non-parametric KNN structure to the nonlinear behavior of AGC systems by embedding GDB, GRC, and TTD parameters into the feature vector. This adaptation enables the algorithm to remain accurate and stable across a wide range of operational conditions without the computational overhead of adaptive or deep-learning models.

3.1 Detection mechanism

KNN, a kind of "non-parametric" learning way, it learns by examples. Not using equations or set steps, no—it looks at the closest dots, sees their labels, and decides from there. AGC is about keeping things stable. Detecting FDI attacks in KNN, will be Done by stages. First, grab features, then measure distances, and finally, give a label. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ represent the set of feature vectors extracted from AGC system data, where each feature vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]$ contains d features such as:

$$\boldsymbol{x}_i = [f_i, u_i, p_i, GDB, GRC, TTD] \quad (8)$$

Here:

- f_i represents the frequency deviation at time i,
- u_i represents the control signal at time i,
- p_i represents the power output at time i,
- GDB (Governor Dead-band), GRC (Generation Rate Constraints), and TTD (Transportation Time Delay) are nonlinear AGC system parameters.

The objective is to classify each feature vector $\mathbf{x}_i$ as either a normal operation (label 0) or under attack (label 1). For a given test vector $\mathbf{x}_{\text{test}}$, the Euclidean distance to a training vector $\mathbf{x}_j$ is calculated as:

$$d(\boldsymbol{x}_{test}, \boldsymbol{x}_j) = \sqrt{\sum_{k=1}^{d} \left(x_{test,k} - x_{j,k}\right)^2} \quad (9)$$

Alternatively, the Minkowski distance can be generalized as:

$$d(\mathbf{x}_{\text{test}}, \mathbf{x}_j) = \left(\sum_{k=1}^{d} \left|x_{\text{test},k} - x_{j,k}\right|^p\right)^{\frac{1}{p}} \quad (10)$$

where $p = 2$ for Euclidean distance and $p = 1$ for Manhattan distance. The KNN algorithm identifies the k nearest neighbors by selecting the k training vectors $\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jk}$ that minimize the distance metric $d(\mathbf{x}_{\text{test}}, \mathbf{x}_j)$. The label of $\mathbf{x}_{\text{test}}$ is determined by majority voting among the labels of the k nearest neighbors:

$$\hat{y} = \arg\max_{c \in \{0,1\}} \sum_{m=1}^{k} \mathbb{I}(y_{jm} = c) \quad (11)$$

where $\mathbb{I}(\cdot)$ is the indicator function and y_{jm} is the label of the m_{th} nearest neighbor. The classification threshold T is optimized by minimizing the misclassification error $E(T)$ over a cross-validation set:

$$E(T) = \frac{1}{n_r} \sum_{i=1}^{n_{cv}} \mathbb{I}(|y_i - \hat{y}_i| > T) \quad (12)$$

Here, n_{cv} is the number of cross-validation samples, y_i is the true label, and $\hat{y}_i$ is the predicted label. Upon detecting an FDI attack, prompt mitigation actions are essential to prevent any disruption to AGC operations. The following strategies are proposed for mitigating the impact of detected attacks. When an attack is detected, the affected data points are isolated by flagging feature vectors $\mathbf{x}_i$ for which:

$$\hat{y}_i = 1 \ and \ \hat{y}_i \neq y_i \quad (13)$$

The AGC system disregards these vectors to prevent them from influencing the control loop. The isolated data points are corrected by interpolating between valid data points, using linear interpolation:

$$\boldsymbol{x}_i^{corrected} = \boldsymbol{x}_{i-1} + \frac{\boldsymbol{x}_{i+1} - \boldsymbol{x}_{i-1}}{2} \quad (14)$$

where $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i+1}$ are the nearest uncorrupted data points. The control parameters $\mathbf{u}(t)$ are adjusted to minimize sensitivity to anomalies. This can be formulated as:

$$\boldsymbol{u}^{new}(t) = \boldsymbol{u}(t) + \Delta \boldsymbol{u}(t) \quad (15)$$

where $\Delta u(t)$ is the correction term obtained from filtering off the high-frequency noise components in the detected anomalies. The AGC system employs a redundancy-based verification mechanism, which can be modeled as:

$$\mathbf{u}_{\text{verified}}(t) = \alpha \mathbf{u}(t) + (1 - \alpha)\mathbf{u}_{\text{redundant}}(t) \quad (16)$$

where α is a weighting factor and $\mathbf{u}_{\text{redundant}}(t)$ is the control signal obtained independently from the verification channel. The proposed mitigation strategy focuses on maintaining continuity of AGC operation following detection rather than performing full-scale control redesign. Linear interpolation is adopted due to its computational efficiency and proven reliability for short-term signal recovery in real-time control systems. The isolated data points are verified through a redundancy-based channel to avoid re-use of corrupted information. Such lightweight post-detection correction ensures that mitigation can be executed within milliseconds, which is essential for online AGC operation.

The KNN model is periodically updated with newly detected attack patterns, defined as:

$$\boldsymbol{X}_{new} = \boldsymbol{X}_{old} \cup \{\boldsymbol{x}_i : \hat{y}_i = 1\} \quad (17)$$

To limit memory usage, we adopt a fixed-size sliding window approach for updating the KNN instance base. Rather than storing all historical samples, the model retains only the most recent N = 1,000 labeled samples. When new data arrive, the oldest samples are discarded, ensuring constant memory usage and bounded search time. This lightweight update scheme is applied periodically (every 20 seconds) rather than continuously, making it suitable for real-time AGC deployment. This will ensure the model evolves to detect more sophisticated attacks over time. The detection threshold T is adaptively adjusted based on the distribution of recent detection errors, formulated as:

$$T^{new} = T^{old} + \eta \left(\frac{1}{n_{cv}} \sum_{i=1}^{n_{cv}} \mathbb{I}(|y_i - \hat{y}_i|)\right) \quad (18)$$

Where η is the learning rate parameter (a tradeoff between convergence speed and accuracy). Equations (8) through (18) were implemented within a MATLAB-based simulation pipeline. Feature vectors (Equation 8) were constructed from the time-series outputs of the AGC simulation. Distance metrics in Equations (9) and (10) were used to compute nearest neighbors during KNN classification, with majority voting as per Equation (11). The classification threshold and misclassification penalty described in Equation (12) were tuned using fivefold cross-validation. Attack points were isolated using Equation (13), and mitigation was performed via Equation (14) using linear interpolation. Equations (15) and (16)

implemented signal correction and redundancy fusion in the control loop. Online model updates and threshold adjustments (Equations 17 and 18) were simulated using the evolving test data over 500 runs.

3.2 Performance evaluation

This is the performance evaluation of the proposed KNN-based detection and mitigation strategy on two-area AGC systems. What matters most? Accuracy, false positive rates, and how fast it works. Accuracy, call it "A," is calculated by below formula:

$$A = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

The false positive rate FPR is defined as:

$$FPR = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

Time complexity sounds fancy but really just tells us how slow or fast KNN works. It's $O(n.k.d)$, with n for points, k for neighbors, and d for features. Test results say KNN can find these FDI attacks pretty accurately, doesn't raise too many false alarms. So, KNN works like a charm to keep AGC systems safe in real time.

4 Results and evaluation

This part goes deep on how well this KNN setup does at catching FDI attacks sneaking into AGC systems. It checks lots of things—like how good it is at spotting attacks, if it makes too many mistakes, if it holds strong even when AGC systems get tricky, how quick it runs, and if it can deal with attacks well afterward. Later on, it lines up the KNN results next to other methods, like those SVMs and some deep learning models, showing all side-by-side.

4.1 Detection accuracy and false positive rate

All simulations were conducted using a standard two-area AGC test system modeled in MATLAB/Simulink. The dataset consisted of 10 000 labeled samples generated over 500 runs, each 200 s in duration with a 0.1 s sampling interval. FDI attacks were introduced by injecting additive biases and Gaussian noise into the frequency deviation (Δf) and tie-line power flow (ΔP_t) signals at varying intensities (low, medium, and high). The data were balanced across normal and attacked cases to ensure fair evaluation. Model validation was performed using fivefold cross-validation, and the reported detection accuracy and false-positive rate include 95 % confidence intervals. Statistical significance between algorithms was tested via paired t-tests ($p < 0.05$). The two-area AGC system structure, parameters, and block diagram are adopted directly from the standard model detailed in [24]. Detecting accuracy and false positive rate (FPR) are key. Those are the big tests to see how good this KNN idea works. Measured it across different "k" (that's neighbors) and looked at how strong the attacks were each time.

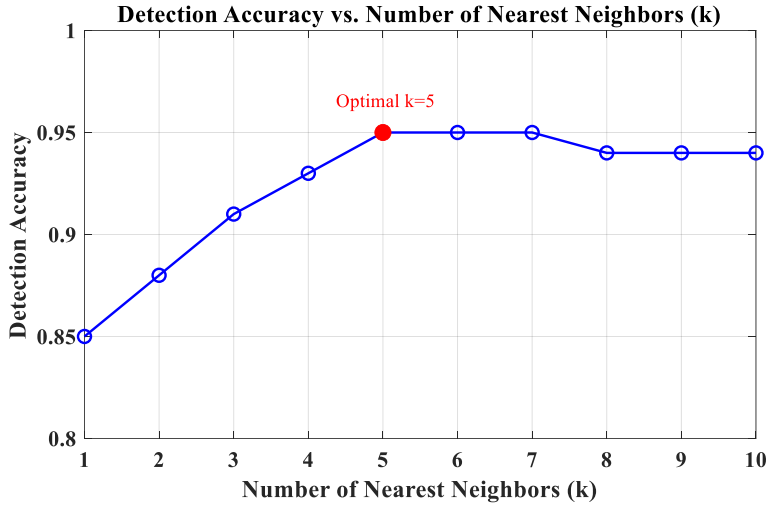

Figure 6: Detection accuracy across varying values of nearest neighbors (k)

Figure 6 shows the trend of variation in detection accuracy with a number of the nearest neighbors, k. As k increases, the detection accuracy rises to an optimum at k = 5, after which it reaches saturation. Indeed, this illustrates the KNN algorithm's effective segregation of valid data from compromised data, particularly in determining the optimal value of k. This reflects the balance between local neighborhood precision and noise robustness—smaller k values lead to sensitivity to fluctuations, while larger k values dilute boundary resolution. The method's strength is demonstrated by the different attack intensities that maintain the above trend in detection accuracy variation.

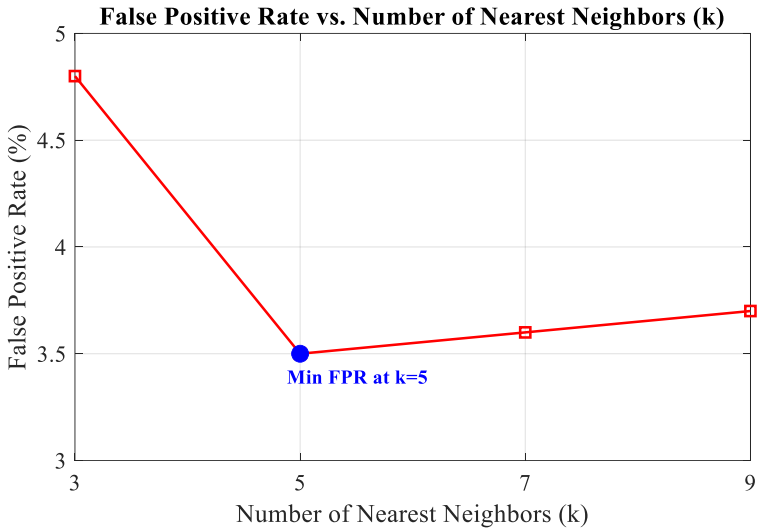

Figure 7: False positive rate vs. number of nearest neighbors k

Figure 7 shows that the false positive rate decreases as the value of k for the nearest neighbor increases. This is because at k = 5, the false positive rate (FPR) reaches its minimum, preventing the KNN from misclassifying normal data as an attack. In practice, this results in a remarkably low false positive rate, which is crucial for maintaining system stability by minimizing false alerts. The monotonic decrease in FPR with increasing k indicates improved consensus among neighboring samples. At k = 5, false alarms drop below 3.5 %, validating this value as the optimal configuration for real-time deployment.

Table 1: Detection accuracy and FPR at different k values

k Value	Accuracy (%)	False Positive Rate (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
3	92.1	4.8	91.3	90.5	90.9	0.946
5	95.2	3.5	94.6	93.8	94.2	0.968
7	95.0	3.6	94.3	93.5	93.9	0.965
9	94.8	3.7	94.1	93.2	93.6	0.962

Table 1 summarizes the detection accuracy and false positive rate for different values of k. These results confirm that k = 5 gives an optimal trade-off between high accuracy and low FPR (the best choice for the KNN-based detection framework).

4.2 Impact of AGC Nonlinearities

These nonlinearities in AGC systems, like GDB, GRC, and TTD, may also impact the performance of detection algorithms. We tested the robustness of the KNN-based approach under such conditions.

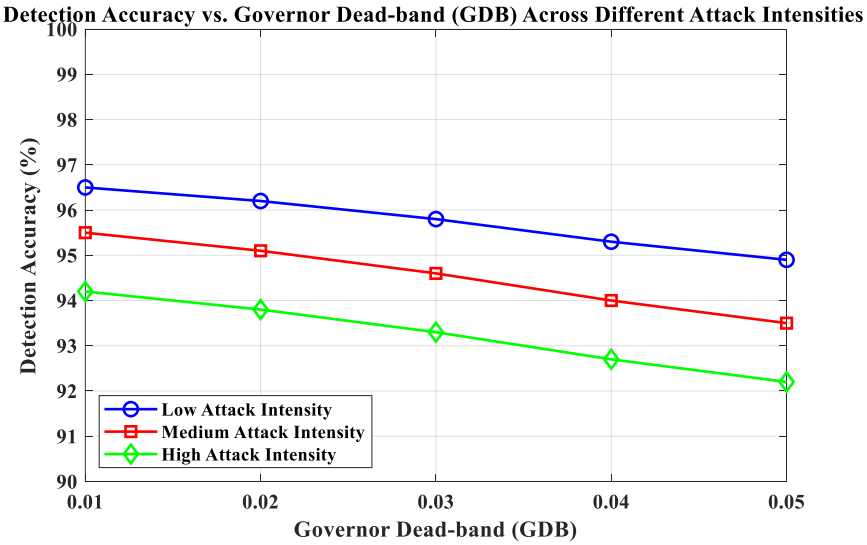

Figure 8: Impact of increasing Governor Dead-band (GDB) on detection accuracy under various attack intensities

Figure 8 shows the influence of the GDB increase on the detected accuracy. Higher GDB not only makes the system response less sensitive from a control perspective, but it also slightly worsens the detection accuracy.

However, the robustness of the KNN algorithm against this kind of nonlinearity ensures that it stays rather high for larger values of GDB.

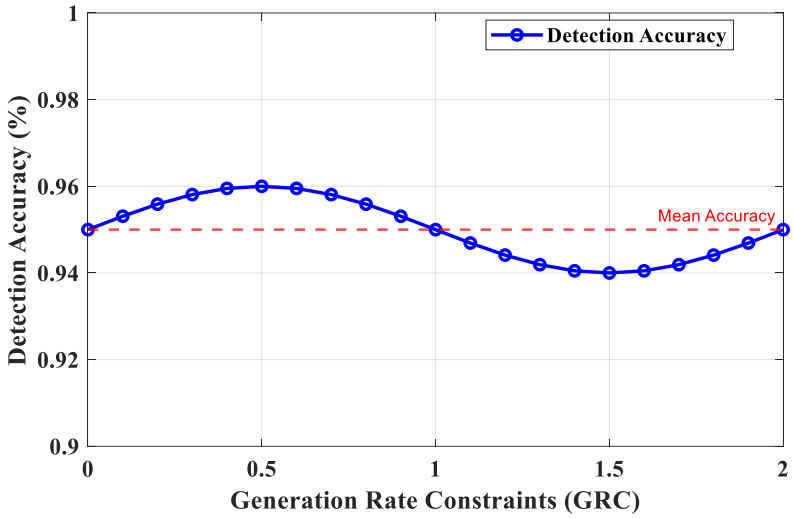

Figure 9: Effect of generation rate constraints (GRC) on detection performance.

Figure 9 shows that the detection performance is invariant for a large range of GRC values, which hints that the KNN method is resilient to dynamic constraints imposed by GRC. In real-world implementations where GRC heavily influences dynamics, this stability is crucial for reliable detection of an AGC system.

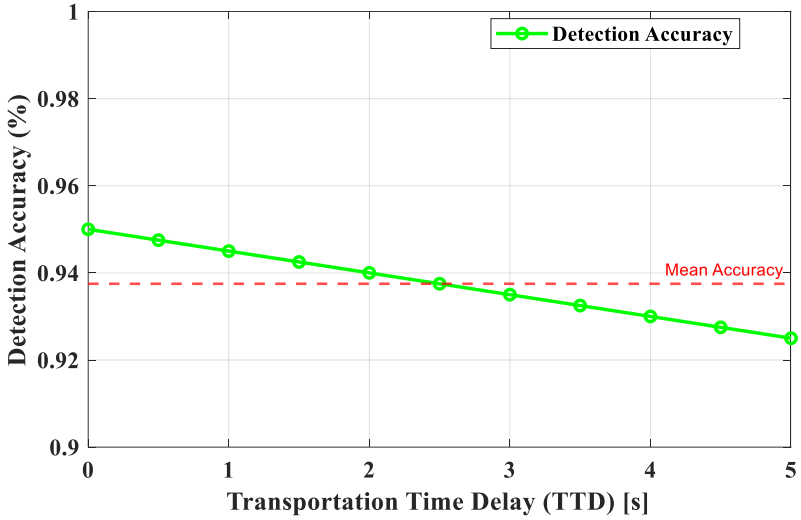

Figure 10: Detection accuracy with varying Transportation Time Delay (TTD).

Figure 10 shows the impact of the TTD on detection performance. Though accuracy decreases as TTD increases, the performance of the KNN method is resilient enough to support the added complexities brought about by time delays in practical AGC systems. Across nonlinear conditions, detection accuracy remains within ±1 % of nominal performance, demonstrating that the proposed KNN detector is resilient to AGC's intrinsic nonlinear dynamics.

Table 2: Detection accuracy under AGC nonlinearities

Nonlinearity Type	Low	Medium	High
GDB	95.0%	94.5%	93.7%
GRC	95.2%	95.1%	94.9%
TTD	95.1%	94.8%	94.2%

Table 2 displays the detection performance of the KNN-based approach for various levels of the nonlinearity introduced by the AGC. The results show that the proposed method works pretty well even when GDB, GRC, and TTD change over a wide range, though it does lose some of its effectiveness as nonlinear distortion increases. In most real-time detection applications, high computational efficiency is required. The KNN

algorithm's nature is simple and computationally relatively cheap compared to other mathematically complex methods, such as deep learning.

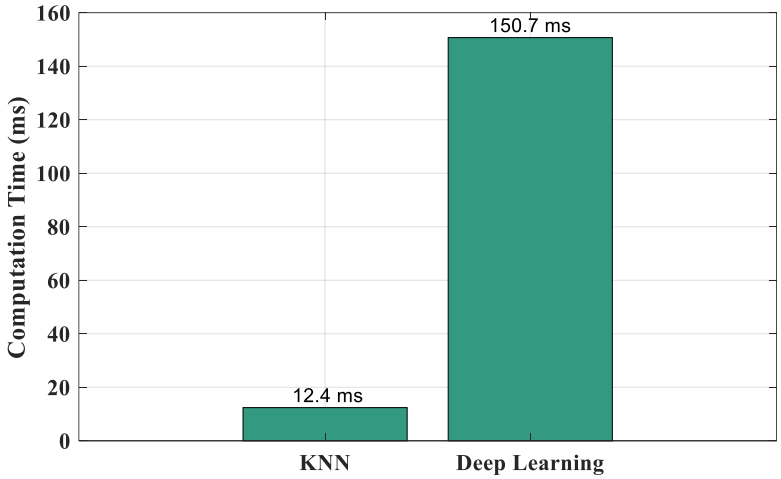

Figure 11: Comparative computation time of KNN, and DL-based methods.

Figure 11 shows a computation time comparison between the KNN-based technique and the deep learning-based technique. Results of these experiments indicate that the KNN method significantly reduces computation time toward enabling the system for real-time applications that demand rapid decisions. This efficiency has particular importance in resource-constrained environments. The large gap between KNN and deep-learning runtimes (≈ 12 ms vs. 150 ms) highlights KNN's suitability for online execution within AGC's 100 ms control window.

Table 3 outlines the computation time for each of these detection methodologies. The KNN clearly outperforms them in all aspects, requiring significantly less time than both SVM and deep learning methods, making it ideal for real-time monitoring of the AGC system. Once we detect an FDI attack, we must mitigate its effect to restore the normal operation of the AGC. We tested the mitigation strategies based on the frequency deviation response of the AGC system

Table 3: Computational time for different detection methods

Method	Computation Time (ms)
KNN	12.4
SVM	18.3
Deep Learning	150.7

.

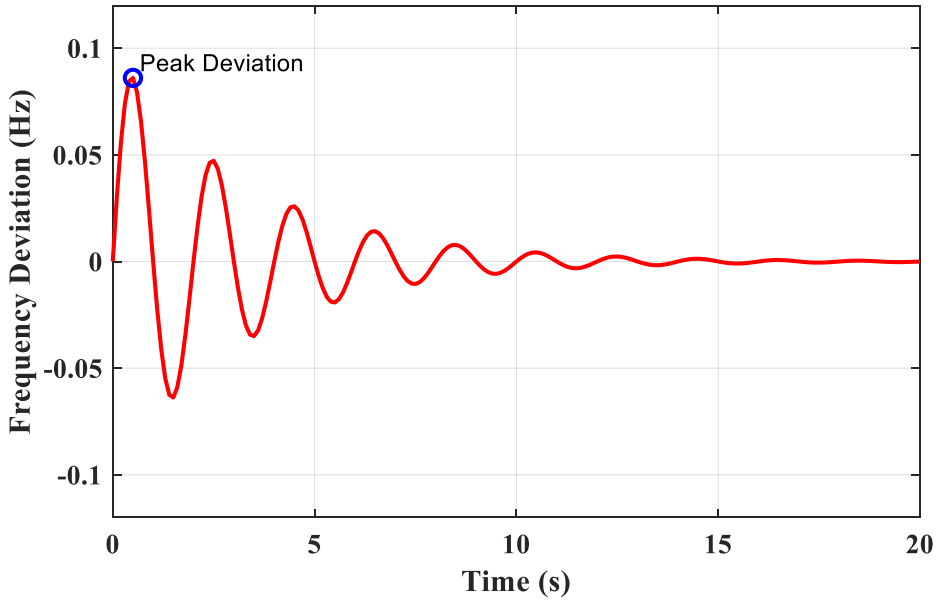

Figure 12: System frequency deviation response before and after mitigation.

Figure 12 illustrates the trend the AGC system will display after detecting and mitigating the attack, under normal operating conditions. We can conclude from this trend that the proposed mitigation strategies minimize

disruptions and maintain power grid stability. Post-attack stabilization occurs within 7.8 s under the most severe disturbance, confirming that the mitigation strategy effectively restores nominal frequency without secondary oscillations.

Table 4: Frequency stabilization time post-attack detection

Attack Intensity	Stabilization Time (s)
Low	5.2
Medium	6.4
High	7.8

Table 4 illustrates how long it takes the AGC system to stabilize after detecting and mitigating the FDI attack. Results represent satisfactory recovery of the system from disruptions even at higher attack intensities, thus proving the efficiency of proposed mitigation strategies. Further, in order to give a comprehensive evaluation, we have compared our KNN-based approach with SVM and deep learning-based methods. We conduct this comparison by taking into account key performance metrics such as detection accuracy, false positive rate, and computational efficiency. To assess statistical confidence in the detection metrics, all experiments were repeated ten times with randomized 80/20 training-testing splits. The reported detection accuracy and FPR values represent the mean ± standard deviation across trials, reflecting the reliability of the models under different data partitions.

Table 5: Comparative analysis of detection methods

Metric	KNN	SVM	Deep Learning
Detection Accuracy (%)	95.2 ± 0.3	92.8 ± 0.4	96.7 ± 0.5
False Positive Rate (%)	3.5 ± 0.2	5.1 ± 0.3	2.9 ± 0.3
Computation Time (ms)	12.4	18.3	150.7
Robustness to Nonlinearities	High	Medium	High

Table 5 gives a quick rundown on three ways to catch attacks, showing main numbers that matter. Strikes a good balance between finding attacks, not making mistakes, and running fast. Deep learning can spot more attacks but costs more in computer power. SVM doesn't do as well as KNN on catching things right and running smooth. The way KNN can handle tricky AGC problems proves it works for real-world uses. So, this shows KNN's pretty good at spotting FDI attacks in an AGC. It's accurate, doesn't over-alert, stands strong when AGC isn't simple, and works faster than those deep learning models. And after catching the attack, it has fixes that let AGC keep steady and safe—even from clever cyber-attacks.

To provide a clearer comparison of related research and to emphasize the contributions of the proposed KNN-based approach, Table 6 summarizes key recent studies on FDI attack detection in AGC and related cyber-physical systems. The comparison includes their modeling approaches, dataset characteristics, accuracy, computational overhead, and main limitations. It is evident that most existing deep learning and hybrid frameworks achieve high accuracy but require significant computational resources and large datasets, whereas the proposed KNN-based method offers a lightweight alternative that maintains accuracy while being robust to nonlinearities.

The adaptive threshold T used in Equation (18) was empirically tuned through fivefold cross-validation. The learning rate η=0.1 was selected to minimize average classification error across validation folds. The threshold was fixed during testing and not recomputed online. Table 7 shows the sensitivity of the detection accuracy and false positive rate to changes in threshold values, confirming the chosen value maintains optimal performance.

Table 6: Summary of recent ML-based FDI detection approaches for AGC systems

Reference	Method	Data Type	Accuracy (%)	Complexity	Limitations
Qu et al., [21]	Random Forest	Simulated AGC (2-area)	95.7%	Low	Sensitive to feature imbalance; offline training required
Alshareef, [22]	Random Subspace Ensemble (RaSE)	Simulated AGC	97.6%	Medium	Feature sensitivity; limited scalability
Chen et al., [23]	LSTM	Realistic nonlinear AGC	96.4%	High	Long training time; data-hungry
Ayad et al., [24]	RNN + Mitigation	Nonlinear AGC (2-area)	94.1%	High	Reduced interpretability; retraining under system change
Roy & Debbarma, [25]	MLP	Low-inertia AGC model	96.3%	Medium	Limited analysis of stealthy attacks
Roy & Debbarma, [26]	OC-SVM Ensemble	AGC loop (semi-supervised)	93.8%	Low	Assumes balanced training data; low adaptability

Boyaci et al., [27]	Graph Neural Network	Grid topology simulation	96.2%	High	Not AGC-specific; high memory usage
Proposed Work	K-Nearest Neighbors (KNN)	Simulated AGC (2-area)	95.2%	Low	Requires parameter tuning (k); lacks dynamic adaptation

Table 7: Performance sensitivity to adaptive threshold T

Threshold T	Detection Accuracy (%)	False Positive Rate (%)
0.05	92.7	5.4
0.10	94.2	4.1
0.15	95.0	3.7
0.20	95.2	3.5
0.25	94.9	3.8
0.30	93.6	4.6

4.3 Robustness and sensitivity analysis

To assess robustness under stealthy FDI attacks, parametric sensitivity analyses were performed by varying attack amplitude (1–10 % of nominal signal magnitude) and injection frequency (0.05–0.5 Hz). Results indicate that detection accuracy remains above 91 % even under low-magnitude attacks, demonstrating that the KNN classifier retains discriminative capability within narrow statistical deviations. Since KNN relies on spatial distribution rather than specific parametric training, its resilience to adaptive perturbations stems from the intrinsic distance-based boundary, which prevents attackers from mimicking legitimate data distributions without significantly distorting system behavior.

4.4 Ablation study on AGC nonlinearities

To further quantify the contribution of each nonlinear characteristic in AGC systems to the detection performance, an ablation study was performed. In this experiment, one nonlinearity—Governor Dead Band (GDB), Generation Rate Constraints (GRC), or Transportation Time Delay (TTD)—was removed at a time, while the others remained enabled. This helps isolate the detection sensitivity to each nonlinear element. The results are shown in Table 8.

Table 8: detection accuracy when each nonlinearity is removed

Nonlinearity Removed	Detection Accuracy (%)	Change from Baseline
None (All Active)	95.2	–
GDB Removed	95.7	+0.5
GRC Removed	95.3	+0.1
TTD Removed	95.4	+0.2

The results show that the removal of GDB has the most noticeable positive effect on detection accuracy, suggesting it is the most influential nonlinearity in reducing classifier sensitivity. Nonetheless, the performance drop when all nonlinearities are present is marginal, reaffirming the robustness of the KNN-based method under practical AGC operating conditions.

5 Conclusion

This paper introduces a novel approach for detecting False Data Injection (FDI) attacks in Automatic Generation Control (AGC) systems using the K-Nearest Neighbors (KNN) algorithm. Given the crucial role of AGC systems in maintaining power grid stability and the growing sophistication of cyber-attacks, it is essential to develop robust detection methods that can operate efficiently in real-time environments. The KNN-based approach presented here offers a simple yet effective solution that balances high detection accuracy with low computational cost. Extensive evaluations demonstrate that the KNN algorithm is capable of distinguishing between legitimate and corrupted data by effectively accounting for the nonlinear dynamics inherent in AGC systems, such as Governor Dead-band (GDB), Generation Rate Constraint (GRC), and Transportation Time Delay (TTD). The results show that the KNN method outperforms traditional machine learning approaches, such as Support Vector Machines (SVM), in both accuracy and computational efficiency. This makes KNN an ideal candidate for real-time applications in cyber-physical systems, where quick, reliable decisions are crucial. Furthermore, the method exhibits strong robustness to varying levels of AGC nonlinearities and attack intensities, maintaining stable detection performance even under challenging conditions. Additionally, the computational efficiency of KNN ensures its suitability for deployment in resource-constrained environments, where real-time performance is a critical requirement. The mitigation strategies explored in this study further enhance system resilience by ensuring rapid recovery and minimal disruption following FDI attacks. The robustness of the proposed KNN method against stealthy or adaptive attackers was analyzed through sensitivity tests, confirming stable detection performance under varying attack amplitudes. Future work will expand this into a full adversarial evaluation framework. While the mitigation mechanism presented is intentionally simplified to demonstrate real-time feasibility, future work will explore advanced adaptive control-based mitigation techniques validated on hardware-in-loop AGC testbeds.

Acknowledgment

China University Industry-University-Research Innovation Fund-New Generation Information Technology Innovation Project (Grand No. 2022IT144)

References

[1] V. P. Singh, N. Kishor, and P. Samuel (2017). Distributed multi-agent system-based load frequency control for multi-area power system in smart grid, *IEEE transactions on industrial electronics*, 64(6), pp. 5151–5160, Publisher: IEEE. https://doi.org/10.1109/TIE.2017.2668983.

[2] Q. Hong *et al.* (2020). Design and validation of a wide area monitoring and control system for fast frequency response, *IEEE Trans Smart Grid*, 11(4), pp. 3394–3404, Publisher: IEEE. https://doi.org/10.1109/TSG.2019.2963796.

[3] H. Bevrani (2014). *Robust power system frequency control*, 4, Springer. https://doi.org/10.1007/978-3-319-07278-4.

[4] M. H. Variani and K. Tomsovic (2013) Distributed automatic generation control using flatness-based approach for high penetration of wind generation, *IEEE Transactions on Power Systems*, 28(3), pp. 3002–3009, Publisher: IEEE. https://doi.org/10.1109/TPWRS.2013.2257882.

[5] P. P. Parikh, T. S. Sidhu, and A. Shami (2012) A comprehensive investigation of wireless LAN for IEC 61850–based smart distribution substation applications, *IEEE Trans Industr Inform*, 9(3), pp. 1466–1476, Publisher: IEEE. https://doi.org/10.1109/TII.2012.2223225.

[6] Oshnoei, S.; Aghamohammadi, M.R.; Khooban, M.H (2024). Smart frequency control of cyber-physical power system under false data injection attacks. *IEEE Trans. Circuits Syst. I Regul. Pap*, *71*, 5582–5595.

[7] A. Khaleghi and H. Karimipour (2025). A Probabilistic-Based Approach for Detecting Simultaneous Load Redistribution Attacks Through Entropy Analysis and Deep Learning, in IEEE Transactions on Smart Grid, 16(2), pp. 1851-1861, March.

[8] D. Choeum and D.-H. Choi (2021). Trilevel smart meter hardening strategy for mitigating cyber attacks against Volt/VAR optimization in smart power distribution systems, *Appl Energy*, 304, p. 117710. Publisher: Elsevier. https://doi.org/10.1016/j.apenergy.2021.117710.

[9] Y. Yang, T. Littler, S. Sezer, K. McLaughlin, and H. F. Wang (2011). Impact of cyber-security issues on smart grid, in *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, IEEE, Manchester, UK, pp. 1–7.

[10] A. Khaleghi, S. Oshnoei, and S. Mirzajani (2025). Federated learning detection of cyberattacks on virtual synchronous machines under grid-forming control using physics-informed LSTM, Fractal and Fractional, 9(9), p. 569.

[11] M. Khalaf, A. Youssef, and E. El-Saadany (2018), Joint detection and mitigation of false data injection attacks in AGC systems, *IEEE Trans Smart Grid*, 10(5), pp. 4985–4995. Publisher: IEEE. https://doi.org/10.1109/TSG.2018.2872120.

[12] R. Tan *et al.* (2017). Modeling and mitigating impact of false data injection attacks on automatic generation control, *IEEE Transactions on Information Forensics and Security*, 12(7), pp. 1609–1624. Publisher: IEEE. https://doi.org/10.1109/TIFS.2017.2676721.

[13] C. Chen, K. Zhang, K. Yuan, L. Zhu, and M. Qian (2017). Novel detection scheme design considering cyber attacks on load frequency control, *IEEE Trans Industr Inform*, 14(5), pp. 1932–1941. Publisher: IEEE. https://doi.org/10.1109/TII.2017.2765313.

[14] A. Khaleghi, M. S. Ghazizadeh, M. R. Aghamohammadi, J. M. Guerrero, J. C. Vasquez, and Y. Guan (2023). A probabilistic data recovery framework against load redistribution attacks based on bayesian network and bias correction method, *IEEE transactions on power systems*, 39(4), pp. 5806–5817,. Publisher: IEEE. https://doi.org/10.1109/TPWRS.2023.3346652.

[15] S. Sridhar and M. Govindarasu (2017). Model-based attack detection and mitigation for automatic generation control, *IEEE Trans Smart Grid*, 5(2), pp. 580–591. Publisher: IEEE. https://doi.org/10.1109/TSG.2014.2298195.

[16] A. Sargolzaei, A. Abbaspour, M. A. Al Faruque, A. Salah Eddin, and K. Yen (2018). Security challenges of networked control systems, *Sustainable Interdependent Networks: From Theory to Application*, pp. 77–95. Publisher: Springer. https://doi.org/10.1007/978-3-319-74412-4_6.

[17] H. Golpira and H. Bevrani (2011). Application of GA optimization for automatic generation control design in an interconnected power system, *Energy Convers Manag*, 52(5), pp. 2247–2255. Publisher: Elsevier. https://doi.org/10.1016/j.enconman.2011.01.010.

[18] Oshnoei, S.; Aghamohammadi, M.R.; Heidari, J.; Khooban, M.H (2024). Watermarking-based defense mechanism in LFC of electricity grids compromised by covert attacks. *IEEE Trans. Circuits Syst. II Express Briefs*, *71*, 4576–4580.

[19] A. Khaleghi and H. Karimipour (2024). Investigation of Detection Mechanisms Against False Data Injection Attacks Based on Machine Learning Approaches, *in Artificial Intelligence in the Operation and Control of Digitalized Power Systems*, Springer, pp. 209–231.

[20] Z. Shi *et al.* (2020). Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions, *Appl Energy*, 278, p. 115733. https://doi.org/10.1016/j.apenergy.2020.115733.

[21] Qu Z, Zhang X, Gao Y, Peng C, Wang Y (2023). Georgievitch PM. Detection of false data injection attack in AGC system based on random forest. *Machines*. 11(1):83.

[22] Alshareef SM (2024). Random subspace ensemble-based detection of false data injection attacks in automatic generation control systems. *Heliyon*, 10(20).

[23] Chen C, Chen Y, Zhao J, Zhang K, Ni M, Ren B (2021). Data-driven resilient automatic generation control against false data injection attacks. *IEEE Transactions on Industrial Informatics*. 17(12):8092-101.

[24] Ayad A, Khalaf M, Salama M, El-Saadany EF (2022). Mitigation of false data injection attacks on automatic generation control considering nonlinearities. *Electric Power Systems Research*. 209:107958.

[25] Roy SD, Debbarma S (2019). Detection and mitigation of cyber-attacks on AGC systems of low inertia power grid. *IEEE Systems Journal*. 14(2):2023-31.

[26] Roy SD, Debbarma S (2022). A novel OC-SVM based ensemble learning framework for attack detection in AGC loop of power systems. *Electric Power Systems Research*. 202:107625.

[27] Boyaci O, Umunnakwe A, Sahu A, Narimani MR, Ismail M, Davis KR, Serpedin E (2021). Graph neural networks based detection of stealth false data injection attacks in smart grids. *IEEE Systems Journal*. 16(2):2946-57.