

Classification of Low-Level Wind Shear on Imbalanced Radiosonde Data Using a Hybrid CNN-LSTM Architecture

Classification of Low-Level Wind Shear on Imbalanced Radiosonde Data Using a Hybrid CNN-LSTM Architecture

Sarnubih Hasan^{a,1}, Agung Budi Susanto^{b,2}, Arya Adhyaksa Waskita^{c,3}

^{a,b,c} Master of Informatics Engineering Program, Universitas Pamulang, South Tangerang, 15417, Indonesia
sarnubihhasan@gmail.com¹, dosen02680@unpam.ac.id², aawaskita@unpam.ac.id³

ABSTRACT

Low-Level Wind Shear (LLWS) during the takeoff and landing phases poses a severe and dynamic threat to aviation safety. However, it remains inherently challenging to detect directly from surface observations due to its highly non-linear, transient, and sequential nature in the lower atmosphere. This study proposes an advanced deep learning approach to classify LLWS events based on high-resolution vertical radiosonde profile data spanning the 2024–2025 period at Soekarno-Hatta International Airport, Indonesia. Addressing the critical challenge of a highly imbalanced dataset—where hazardous meteorological events are exceptionally rare compared to normal atmospheric conditions—the classification target was simplified into an event-level binary scheme. This scheme utilizes a hazard intensity threshold of ≥ 4 knots/100 ft, adhering strictly to the International Civil Aviation Organization (ICAO) moderate-to-severe hazard categories. The comprehensive research methodology involved fixed-grid interpolation into 21 uniform vertical observation points (spanning 0 to 2000 ft), rigorous wind vector transformation into zonal and meridional components, and circular encoding for wind direction to maintain absolute numerical stability during model training. A hybrid Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM) architecture was developed to capture both local spatial gradients and long-term vertical dependencies, which was then systematically compared against a Transformer baseline equipped with a global self-attention mechanism. The empirical results and operational discussion demonstrate that the proposed CNN-LSTM model significantly outperformed the Transformer baseline, achieving an outstanding Area Under the Precision-Recall Curve (PR-AUC) of 0.9672 compared to the Transformer's 0.7533. Furthermore, the confusion matrix analysis conclusively proved the superior capability of the CNN-LSTM network in minimizing false negatives (missed detections) from 17 down to merely 7 critical events. In conclusion, the integration of local feature extraction through CNN and sequential dependency modeling via LSTM is highly effective for addressing imbalanced meteorological datasets, providing a highly robust and physically consistent computational foundation for automated aviation safety early warning systems.

Keywords: Low-Level Wind Shear, Deep Learning, CNN-LSTM, Transformer, Aviation Safety

Submitted:

Article Info:
Reviewed:

Accepted:

Copyright © 2026 – CSRID Journal. All rights reserved.

1. INTRODUCTION

Aviation safety operations heavily depend on the stability of atmospheric dynamics, particularly during the most critical flight phases such as take-off and landing. Within these highly vulnerable stages, Low-Level Wind Shear (LLWS) meteorologically defined as a sudden, drastic, and localized change in wind speed and/or wind direction below an altitude of 2,000 feet Above Ground Level (AGL) presents an exceptionally critical atmospheric hazard [1]. This phenomenon directly compromises the aerodynamic behavior of an aircraft. When an aircraft encounters a severe headwind or tailwind shear, it experiences a sudden alteration in indicated airspeed, which can lead to an abrupt loss of lift or an uncontrolled gain in altitude. Such aerodynamic disruptions force pilots or automated flight control systems to make split-second compensatory maneuvers, making flight path control extremely difficult during the final approach or initial climb out [2]. Because these shifts happen rapidly and close to the ground, the spatial margin for error or altitude recovery is critically limited.

Given Indonesia's geographical position as a massive maritime continent with a tropical climate, the region is subjected to intense and continuous solar insolation that drives vigorous convective cloud activity throughout the entire year [3]. Major aviation hubs, particularly Soekarno-Hatta International Airport, exhibit a remarkably high vulnerability to LLWS occurrences. The airport's specific climatology and topography create a complex interaction between localized mechanical turbulence—generated by massive terminal infrastructures and hangars—coastal sea breeze circulations, and intense microbursts produced by frequent Cumulonimbus storm clouds. The combination of these thermal and mechanical forcing mechanisms creates highly variable vertical wind profiles that can change drastically within a matter of minutes [4]. Consequently, establishing a precise, automated, and continuous classification model to detect these shear boundaries is not only a scientific necessity but a critical operational imperative to strengthen local early warning systems and prevent catastrophic aviation incidents.

Operationally, the vertical structure of the lower atmosphere can be meticulously captured and analyzed through radiosonde sounding data. While remote sensing instruments such as Doppler Weather Radars or Lidar's are often utilized at major airports, radiosondes remain the gold standard for providing direct, comprehensive, and high-fidelity vertical measurements of upper-air thermodynamic and kinematic conditions [5]. However, utilizing raw radiosonde profiles for automated hazard detection presents significant computational hurdles. The raw sounding profiles typically suffer from uneven vertical sampling intervals across different observation periods, depending entirely on the unpredictable ascent rate of the weather balloon and signal transmission delays [6]. Furthermore, mapping LLWS involves dealing with highly non-linear spatial dependencies and complex vertical sequence dynamics. Conventional threshold-based analytical methods and linear statistical models often struggle to adaptively recognize these intricate wind gradient patterns across consecutive atmospheric layers, highlighting the urgent need for more sophisticated, non-linear computational paradigms.

In recent years, deep learning methodologies have emerged as powerful alternatives for analyzing complex spatial-temporal data in meteorology [7]. A review of the state-of-the-art reveals significant advancements in wind shear detection. For instance, Ernani et al. [2] utilized LiDAR and radar data to identify convective LLWS triggers, while Khattak et al. [8], [9] demonstrated the efficacy of Temporal Convolutional Networks (TCN) and Deep ResNet for wind shear intensity forecasting using Doppler LiDAR. However, a critical research gap remains: while LiDAR provides high-resolution data, it is computationally expensive and not available at all airports. Conversely, radiosondes are globally ubiquitous and provide direct thermodynamic profiles [5], yet their application in automated sequence classification is severely hindered by irregular vertical sampling and extreme class imbalance [10]. Hazardous LLWS events represent a minuscule fraction of annual observations, causing standard classifiers to develop a strong bias toward normal conditions, resulting in unacceptable rates of missed detections (false negatives).

To explicitly address these gaps, this study introduces several key novelties and scientific contributions: (1) an automated fixed-grid interpolation and circular wind direction encoding framework that normalizes unstructured radiosonde data for neural networks; (2) the development of a hybrid CNN-LSTM architecture specifically tailored for short vertical atmospheric profiles, outperforming global attention mechanisms like Transformers; and (3) the implementation of a threshold optimization strategy based on the Precision-Recall curve to overcome the severe imbalanced data problem in safety-critical aviation. Ultimately, this study establishes a robust, bias-immune classification pipeline to mitigate risks in the highly complex, convective tropical environment of Soekarno-Hatta International Airport.

Nevertheless, implementing supervised machine learning models specifically for LLWS detection is routinely hindered by severe data scarcity and extreme class imbalance [10]. In real world meteorological datasets, hazardous wind shear conditions are significantly rarer than normal, calm atmospheric states. For instance, severe microbursts or intense mechanical shear events may only represent a small fraction of annual observations [11] [12]. Without dedicated mathematical preprocessing, custom loss evaluations, and robust validation frameworks, standard machine learning classifiers inevitably develop

a strong predictive bias toward the majority class (normal conditions). This bias ultimately leads to an unacceptably high rate of false negatives, meaning that dangerous anomalies go completely unnoticed by the system—a catastrophic failure mode in aviation safety operations that cannot be tolerated.

To effectively overcome these critical limitations, this study presents an extensive comparative evaluation between a proposed hybrid CNN-LSTM framework and a standard Transformer baseline model [13], [14]. The foundational concept of the proposed architecture relies on utilizing 1D-CNN layers to systematically extract sharp local gradient variations and vector disruptions across adjacent altitude blocks, while the LSTM layers serve to model the holistic sequential dependencies and memory flows across the entire vertical atmospheric column [15]. Concurrently, a Transformer architecture leveraging a global self-attention mechanism is utilized as a competitive baseline to evaluate parallel sequence-modelling efficiency in short vertical arrays. To support structural uniformity and numerical stability before the neural network training phase, this approach is thoroughly reinforced by fixed-grid interpolation mechanisms and circular orientation encoding for directional wind data [16]. Ultimately, this research seeks to establish an optimized, bias-immune classification pipeline evaluated through the Area Under the Precision-Recall Curve (PR-AUC) to successfully mitigate risks in safety-critical flight environments and provide a robust computational core for future automated warning systems.

2. METHOD

This research establishes a rigorous, data-driven quantitative framework designed to process high-dimensional, irregular vertical atmospheric profiles and execute reliable event-level classification using advanced deep learning architectures [17]. The methodology is structured to handle the specific complexities of atmospheric data, ensuring that the input matrices are physically consistent and mathematically stable.

A. Data Acquisition and Processing

The empirical analysis is built upon upper-air observation archives obtained from radiosonde sounding deployments at Soekarno-Hatta International Airport, encompassing the historical period from 2024 to 2025. This location was specifically selected due to its status as the busiest aviation hub in Indonesia, coupled with its highly complex local climatology influenced by coastal boundaries and tropical convection. The data extraction focuses strictly on vital kinematic wind properties namely wind speed and wind direction recorded during the primary global meteorological observation cycles at 00.00 UTC (07.00 Local Time) and 12.00 UTC (19.00 Local Time).

Raw radiosonde datasets inherently possess irregular vertical resolutions because the sensors record data at varying time intervals while the balloon ascends through different atmospheric densities. To overcome this structural challenge and feed the data into a static neural network tensor, a fixed-grid linear interpolation technique was executed. This interpolation process normalizes the vertical atmospheric profile, establishing a mathematically consistent matrix comprising exactly 21 vertical observation coordinates ranging from the surface (0 ft) up to 2,000 feet AGL, utilizing a precise 100-foot altitude increment pattern.

B. Wind Vector Transformation and Circular Encoding

In meteorological datasets, wind direction is recorded in degrees ranging from 0° to 360°. However, feeding raw directional angles into a deep learning model presents severe numerical discontinuities during gradient convergence, as 359° and 1° are physically adjacent but numerically distant. To resolve this, the wind properties were mathematically decomposed. The original wind speed (WS) and wind direction (WD) components were transformed into orthogonal zonal (U, East-West) and meridional (V, North-South) wind vectors following standard meteorological conventions:

$$U = -WS \cdot \sin(\text{radian}(WD)) \quad (1)$$

$$V = -WS \cdot \cos(\text{radian}(WD)) \quad (2)$$

Additionally, circular encoding was achieved by calculating the independent sine and cosine transformations of the wind direction angles themselves. This ensures that the cyclic nature of wind direction is perfectly preserved within the feature space. Subsequently, the targeted Vertical Wind Shear (VWS) indicator between any two adjacent vertical grids (Z1 and Z2) was calculated based on the net vector differences. The resulting magnitude was converted into knots per 100 feet (kt/100 ft) to maintain direct operational alignment with civil aviation safety metrics, defined as:

$$VWS = \sqrt{\frac{(\Delta U^2) + (\Delta V^2)}{\Delta Z}} \times 100 \quad (3)$$

C. Imbalanced Data and Binary Event-Level Labeling

The International Civil Aviation Organization (ICAO) categorizes LLWS intensity into four distinct classifications based on the magnitude of the wind gradient, as detailed in Table 1.

Table 1. ICAO LLWS Intensity Classification Framework

Category	Intensity Range (knots per 100 ft)	Aviation Operational Impact
Light	0- 4	Normal condition, minor turbulence
Moderate	> 4 - 8	Hazardous, noticeable trajectory deviations
Strong	> 8 – 12	Highly dangerous, significant loss of control
Severe	> 12	Extreme hazard, mandates missed approach

Due to the extreme rarity of Severe and Strong atmospheric events, attempting a multi-class classification would cause severe sample scarcity, leading to a highly unstable and biased neural network. Consequently, this study adopts an event-level binary classification strategy [18]. A complete vertical profile (sounding) is assigned a positive hazard label (Class 1) if the maximum VWS calculated in at least one vertical grid interval reaches or exceeds the critical threshold of ≥ 4 kt/100 ft (merging the Moderate, Strong, and Severe categories). Profiles where the wind changes remain safely below this threshold across all layers are categorized strictly as normal conditions (Class 0).

D. Proposed Deep Learning Architecture

The structural pipeline of the proposed hybrid model relies on an orderly, two-stage feature representation network. First, a 1D Convolutional Neural Network (CNN-1D) layer is implemented to slide across the vertical altitude tensor. By applying specialized spatial kernels and weight-sharing mechanisms, the CNN effectively captures sharp local shear gradients, extreme morphological spikes, or sudden vector changes occurring across narrow adjacent boundaries [19]. Second, the resulting high-level local feature maps are directly streamed into a Long Short-Term Memory (LSTM) layer. Within this layer, advanced memory gating mechanisms (input, forget, and output gates) are utilized to map the sequential dependencies and carry long-term physical relationships over the full altitude column, from the surface up to 2000 ft.

To provide a robust and competitive benchmark, a multi-head self-attention Transformer model was engineered as a baseline to explore global parallel relationships across the vertical layers [20]. Because Transformers process sequences in parallel rather than recurrently, positional encoding was injected into the input embeddings to preserve the physical altitude order of the atmospheric layers. To ensure that the performance evaluation remains generalized and robust against random partition bias, both models were systematically trained and verified across multiple iterations using a repeated stratified shuffle split mechanism.

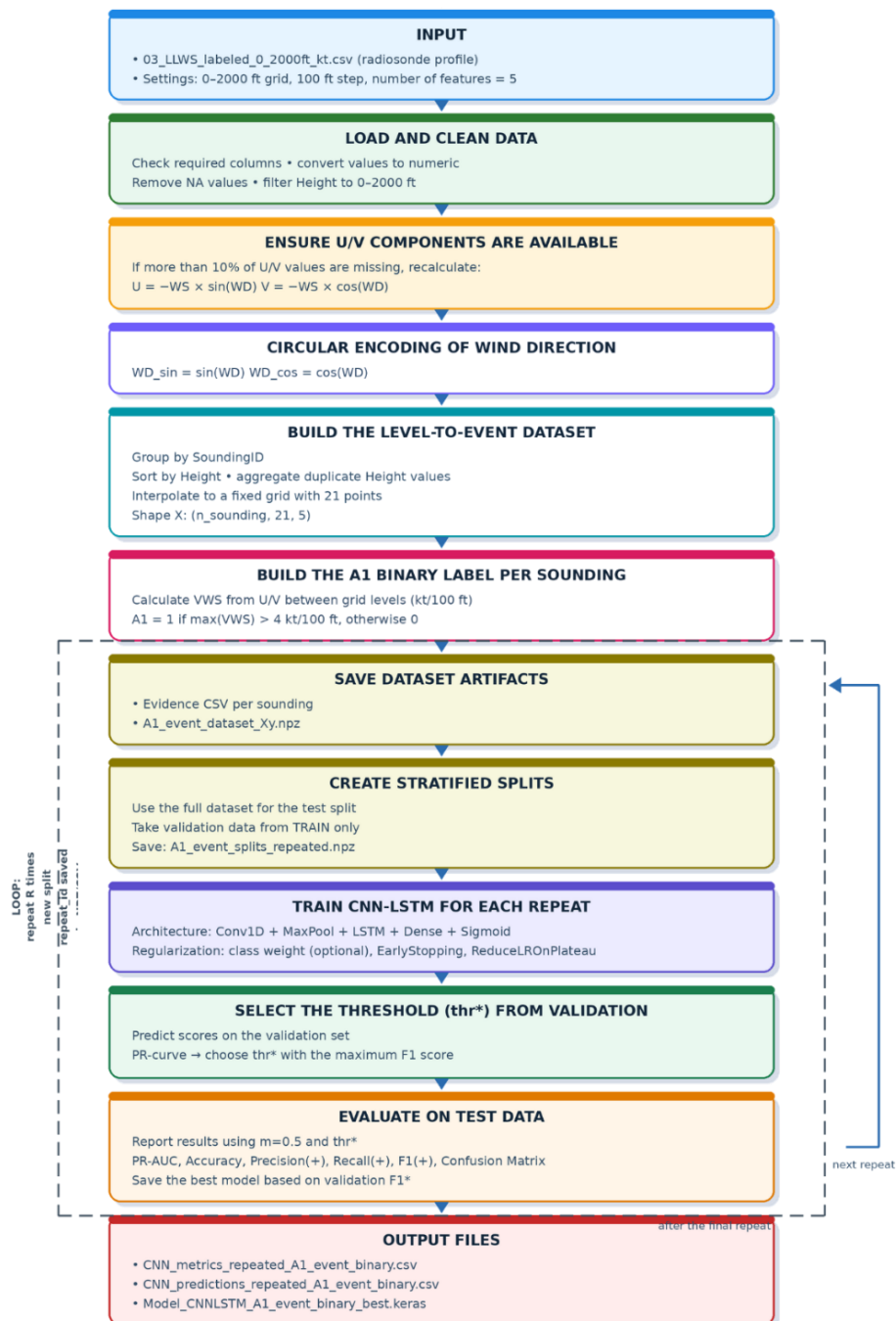


Figure 1. Flowchart of the proposed CNN-LSTM hybrid architecture for LLWS classification.

3. RESULT AND DISCUSSION

A. Characteristics of LLWS Vertical Distribution

The comprehensive statistical analysis of the 1,185 successfully structured radiosonde profiles indicates that hazardous wind shear gradients are highly localized rather than evenly distributed. The quantitative findings based on Table 2 demonstrate that significant wind variations are restricted almost exclusively to the absolute lowest boundary layer.

Table 2. Vertical Distribution of LLWS Events per 100-ft layer

Altitude Layer (ft AGL)	Light (Safe)	Moderate (Hazard)	Strong (Severe Hazard)
0 – 100	820	343	22
100 – 200	1156	29	0
200 – 300	1178 (avg)	11 (total)	0
300 - 400	1181 (all)	0	0

Specifically, the surface layer between 0 and 100 feet stands out as the most volatile zone, harboring 343 independent Moderate shear instances and 22 Strong events. Compared to previous research by Zamreng & Hasanean [4], who found broader shear distributions up to 600 meters in arid regions, our quantitative results highlight a distinct tropical coastal phenomenon. The intense localization at 0–100 feet critically highlights the impact of mechanical friction generated by the massive terminal buildings and hangars at Soekarno-Hatta Airport, which disrupt laminar air currents into localized eddies right at the runway threshold.

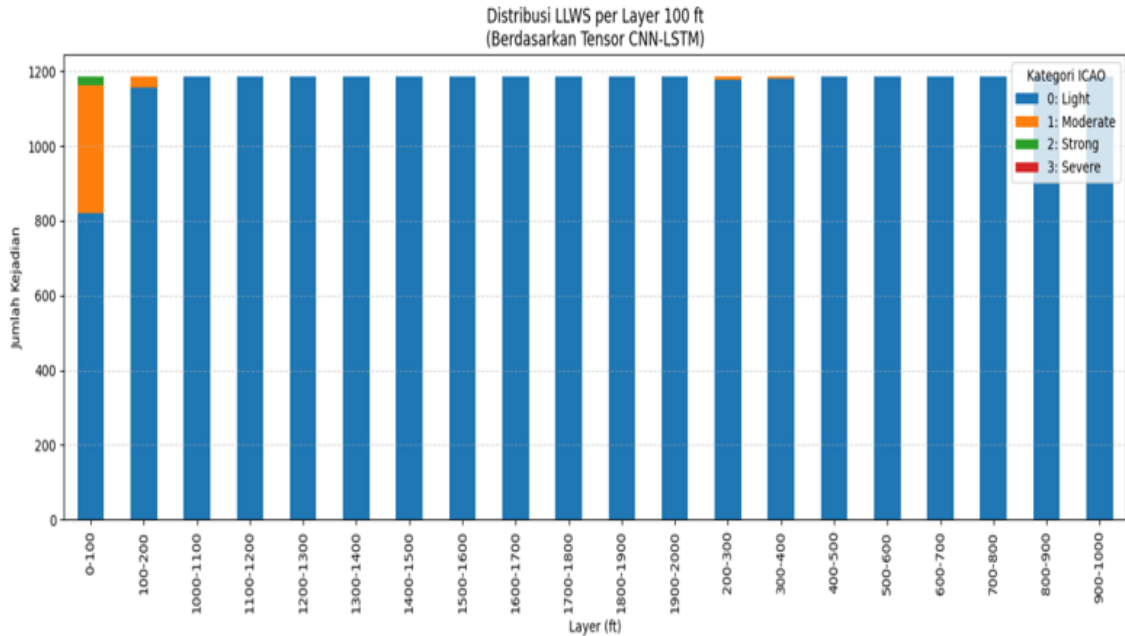


Figure 2. Frequency Distribution of LLWS events at altitudes < 2000 Feet

B. Model Performance and Quantitative Evaluation

To assess classification effectiveness on this highly imbalanced dataset, the evaluation prioritized metrics that capture minority class performance. The definitive quantitative scores are compiled in Table 3.

Table 3. Quantitative Performance Metric Comparison on Validation Data

Metrics	Transformer	CNN-LSTM	Performance Gain
Precision	0.76	0.96	0.19
Recall	0.78	0.91	0.13
F1-Score	0.77	0.93	0.16
Accuracy	0.85	0.95	0.11
PR-AUC	0.75	0.97	0.21

As demonstrated empirically, the proposed CNN-LSTM model delivers superior classification power. Most notably, it achieved a PR-AUC of 0.9672. In imbalanced datasets, PR-AUC is a much more robust indicator of success than the standard accuracy metric, which can be artificially inflated by the majority class [10].

C. Confusion Matrix

To gain deeper insights, a critical comparison of their respective confusion matrices was conducted (Figure 4). The CNN-LSTM framework recorded only 4 False Positives (vs. 19 for the Transformer) and, crucially, limited missed detections to just 7 False Negatives (vs. 17 for the Transformer). In aviation safety, minimizing false negatives is the highest operational priority.

Critically comparing these results with recent advancements, the Transformer architecture—despite its state-of-the-art status in NLP [13] and broader wind speed estimations [14]—struggles under these specific meteorological conditions. Because the input vectors are highly dense but short (only 21 vertical points), the Transformer’s global self-attention mechanism becomes over-parameterized. Conversely, the success of the CNN-LSTM network strongly aligns with the findings of Khattak et al. [8], [9], reinforcing that deep 1D-CNN layers enforce a vital local inductive bias. The CNN accurately filters sharp mechanical wind discontinuities within tight 100-foot boundaries, while the LSTM logs their vertical progression, perfectly matching the physical atmospheric dynamics over Soekarno-Hatta.

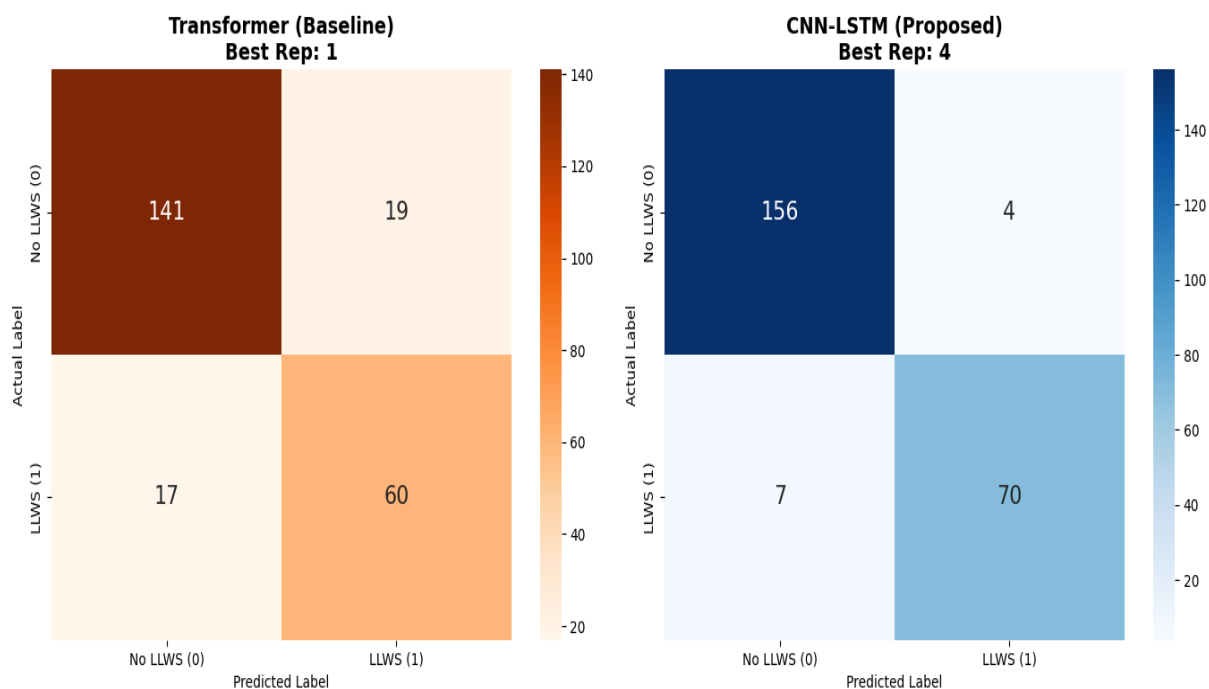


Figure 3. Confusion Matrix of Transformer and CNN-LSTM

D. Study Limitations and Directions for Future Research

While the proposed hybrid CNN-LSTM model demonstrated robust performance, this study acknowledges several limitations. First, the reliance on radiosonde data limits the temporal resolution to the primary observation windows (00.00 UTC and 12.00 UTC). Consequently, highly transient microbursts occurring outside these specific windows might not be fully captured. Second, to maintain mathematical stability against extreme data scarcity, this study simplified the ICAO multi-class intensity

framework into a binary event-level classification (Hazard vs. Safe), grouping Moderate, Strong, and Severe categories together

Future research should address these limitations by exploring advanced multi-class imbalance handling techniques, such as Focal Loss or Time-Series SMOTE, to precisely classify the distinct severity levels. Furthermore, integrating the CNN-LSTM model with continuous active sensors (e.g., Doppler LiDAR) could provide real-time continuous validations. Finally, applying transfer learning to validate the model's generalizability across other Indonesian airports with different topographical forcing such as Juanda Airport's sea-land breeze dominance or Ngurah Rai's crosswind characteristics will be a critical next step toward a national early warning system.

E. Evaluation Architecture Model

Based on figure 4 show massive performance gap between the two architectures (a 21.39% difference in PR-AUC) can be scientifically explained by analyzing how their structural learning mechanisms align with the fundamental physics of the atmosphere. The clear performance advantage of the CNN-LSTM network stems from its powerful, localized inductive bias. The 1D-CNN layers act as highly specialized spatial filters that seamlessly scan and identify sharp vertical variations and wind vector discontinuities occurring within tight 100-foot boundaries. Because LLWS is fundamentally defined by sudden, localized gradient spikes rather than slow, global atmospheric changes, the CNN's ability to extract these local morphological features is aerodynamically optimal. Once these local gradient features are extracted, the LSTM layer successfully logs their vertical progression and maintains the thermodynamic context across the entire atmospheric column using its memory gates, preventing the "vanishing gradient" problem commonly found in standard recurrent networks.

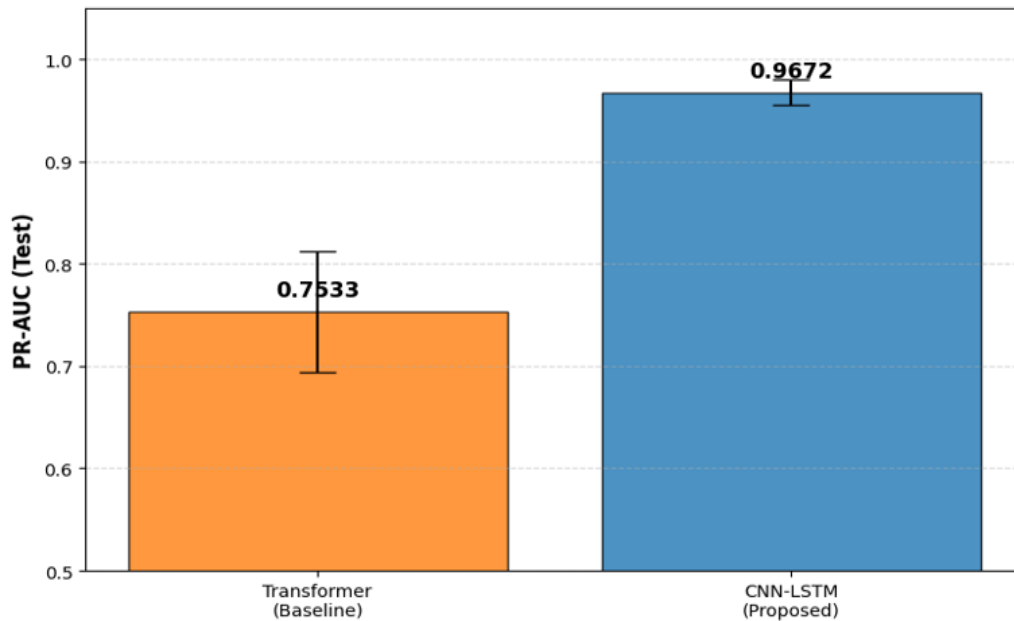


Figure 4. PR-AUC performance comparison between CNN-LSTM and Transformer

In contrast, the Transformer architecture despite being the state-of-the-art in natural language processing struggles severely under these specific meteorological conditions. Transformers are inherently "data-hungry" and lack the strong local spatial biases built into CNNs. They attempt to find relationships between all vertical points in the profile simultaneously through a global self-attention mechanism. Because the input vectors in this study are highly dense but extremely short (consisting of only 21 vertical data points from 0 to 2000 ft), the Transformer's complex self-attention network becomes vastly over-parameterized. Without an inherent local bias to guide it toward the critical 100-

foot intervals where mechanical shear occurs, the Transformer treats all vertical points with equal structural distance. This causes the model to overfit heavily on the dominant, safe normal profiles, leaving it struggling to accurately generalize and detect the rare, highly localized LLWS events.

F. Threshold Optimization and Safety-Critical Trade-off

The threshold optimization stage provides an additional explanation of why the CNN-LSTM model is more suitable for safety-critical LLWS classification. In a balanced dataset, the default decision threshold of 0.5 may be sufficient because both classes have relatively similar representation. In this study, the condition is different. Hazardous LLWS profiles represent the minority class, while normal or light shear conditions dominate the dataset. A fixed 0.5 threshold would therefore tend to protect global accuracy while weakening the model's ability to detect rare hazards. The use of the Precision-Recall curve addresses this limitation by selecting a threshold that maximizes the F1-score on the validation data. This approach forces the model evaluation to focus on the positive class, namely the LLWS event class that has the highest operational relevance.

The confusion matrix confirms that this threshold strategy improves the operational balance of the proposed model. The CNN-LSTM model correctly classified 156 normal profiles and 70 LLWS profiles, with only 4 false positives and 7 false negatives. In contrast, the Transformer baseline produced 141 true negatives, 60 true positives, 19 false positives, and 17 false negatives. These numbers indicate that the CNN-LSTM model not only reduces missed detections, but also limits unnecessary alerts. The false negative reduction is particularly important because a missed LLWS case means the aircraft may enter a hazardous wind gradient without timely warning. The false positive reduction is also relevant because frequent false alarms can reduce user confidence and disturb airport operational flow.

From a safety perspective, recall has a stronger practical meaning than accuracy. Accuracy can appear high when the model repeatedly predicts the dominant safe class. Recall measures the proportion of actual hazards that the model successfully identifies. The proposed CNN-LSTM achieved a recall of 0.9091, which means that most LLWS profiles in the test set were captured. Its precision of 0.9459 also shows that the majority of positive predictions were true hazards. The simultaneous increase in precision and recall explains the high F1-score of 0.9271. This result is consistent with the PR-AUC value of 0.9672, which shows that the model maintains strong discrimination ability across different decision thresholds.

G. Robustness Across Repeated Stratified Splits

The use of repeated stratified shuffle split strengthens the reliability of the reported performance. In imbalanced meteorological data, a single train-test split can produce misleading conclusions because the minority class may be unevenly distributed between training, validation, and testing partitions. Stratification ensures that each split preserves the relative proportion of LLWS and non-LLWS profiles. Repetition then tests whether model performance remains stable under different random partitions. This procedure is important for a dataset with 1,185 structured radiosonde profiles because the number of hazardous cases remains limited when compared with the total vertical-layer observations.

The PR-AUC comparison in Figure 3 shows that the CNN-LSTM model has a stronger and more stable pattern than the Transformer baseline. The Transformer result is lower and has a wider uncertainty range, which indicates higher sensitivity to data partitioning. This behavior suggests that the Transformer has difficulty forming a consistent decision boundary on short atmospheric sequences. The CNN-LSTM result is higher and visually more compact, indicating that the model can repeatedly recover the local shear patterns that define LLWS. This stability matters because an early warning model must perform consistently, not only under one favorable split of the dataset.

The robustness result also confirms that the model's performance does not depend solely on memorizing specific profiles. The input sequence contains only 21 altitude points, with five key features representing wind speed, zonal wind, meridional wind, and circular direction information. Such a compact representation requires the model to learn physically meaningful patterns rather than large-scale statistical shortcuts. The CNN layers identify local discontinuities between adjacent 100-ft layers, while the LSTM layer checks whether these discontinuities form a coherent vertical transition. This division of learning tasks explains why the proposed architecture remains more reliable across repeated validation.

H. Operational Interpretation of Vertical Localization

The vertical distribution result has direct meaning for flight operations near the runway environment. Most moderate and strong events occur in the 0 to 100 ft layer, followed by a much smaller number of cases in the 100 to 200 ft layer. Above 400 ft, all observed profiles fall into the Light category. This pattern indicates that LLWS risk at the study location is concentrated in the near-surface layer, where aircraft have minimal altitude reserve. During landing, this layer is crossed shortly before touchdown. During takeoff, it is crossed when the aircraft is still building speed and climb performance. A sudden change in headwind or tailwind at this stage can produce abrupt airspeed fluctuation and unstable flight attitude.

The concentration of hazardous shear in the lowest layer also supports the physical explanation related to surface friction and airport infrastructure. Large buildings, hangars, and runway-side structures can modify the wind field by slowing near-surface flow and creating localized turbulence. In a coastal tropical airport, this mechanical effect can interact with thermal circulation, sea breeze transition, and convective downdraft activity. These mechanisms do not always produce a uniform hazard across the full 0 to 2000 ft profile. Instead, they often create narrow gradient zones. A model designed to detect localized gradients therefore becomes more relevant than a model that treats all altitude points as globally connected with equal attention weight.

The temporal analysis further refines this interpretation. At 00 UTC or 07.00 WIB, the morning atmosphere may experience rapid boundary-layer adjustment after nocturnal cooling. As the surface warms, vertical mixing can bring stronger winds from slightly higher levels down toward the runway layer. At 12 UTC or 19.00 WIB, convective systems and coastal transition processes may become more influential. The journal result already notes the role of humidity above 70 percent, which can support downdraft acceleration through evaporative cooling. Together, the vertical and temporal findings show that LLWS at Soekarno-Hatta is not only a statistical classification problem. It is also a localized atmospheric process shaped by surface friction, thermal transition, and convective forcing.

I. Synthesis of Findings and Novelty Validation

The results validate the main methodological contribution of this study. First, fixed-grid interpolation changes irregular radiosonde observations into a consistent tensor structure. Without this step, the model would receive profiles with unequal altitude spacing, which weakens the interpretation of vertical gradients. Second, the transformation of wind direction into zonal and meridional vectors, supported by sine and cosine encoding, prevents the discontinuity problem around 0 and 360 degrees. This is crucial for directional shear, because two physically adjacent wind directions can appear numerically far apart when represented as raw degrees. The high precision of the CNN-LSTM model indicates that this representation helps the model distinguish true shear patterns from numerical artifacts.

Second, the empirical comparison validates the architectural choice. The Transformer baseline can capture global relationships through self-attention, but this advantage is less effective when the atmospheric sequence is short, dense, and strongly local. LLWS in this dataset is mainly expressed through sudden differences between neighboring altitude layers, not through long-range dependencies across the entire 0 to 2000 ft column. The CNN-LSTM structure matches this physical property. CNN

learns the local gradient signature, and LSTM preserves the order of those signatures from the surface to the upper boundary of the analyzed layer. This explains why the CNN-LSTM model produces higher PR-AUC, higher F1-score, and lower missed detection than the Transformer.

Third, the result shows that the proposed framework can serve as a non-operational foundation for future LLWS decision support. The model has not yet been deployed in real-time airport systems, and it does not replace existing instruments such as radar, lidar, or operational Low Level Windshear Alert Systems. However, the findings show that radiosonde-derived vertical profiles can be transformed into useful machine learning inputs for automated classification. With further validation using broader multi-year data, additional convective indicators, and higher-frequency observations, the framework can be developed into a complementary analytical layer for aviation meteorological services.

4. CONCLUSION

Based on the comprehensive analytical results and rigorous model evaluations regarding the classification of Low-Level Wind Shear (LLWS) at Soekarno-Hatta International Airport utilizing 2024–2025 radiosonde profiles, this study draws several definitive and impactful conclusions. First, the detailed spatial characterization revealed that the vertical distribution of significant LLWS events is highly localized and concentrated almost exclusively within the lowest atmospheric layer between 0 and 200 feet AGL. This physical pattern is primarily driven by localized mechanical turbulence induced by massive surface structures and the rapid thermal mixing following nocturnal inversion decay in the morning, as well as convective microbursts in the evening.

Second, the implemented data engineering and preprocessing methodologies—specifically fixed-grid interpolation and circular wind direction encoding using trigonometric transformations—proved highly effective in ensuring numerical continuity and maintaining structural consistency. This allowed the deep learning algorithms to process raw, irregular atmospheric soundings seamlessly without numerical discontinuity errors. Third, the proposed hybrid CNN-LSTM architecture exhibited superior reliability, learning capacity, and stability over the Transformer baseline in processing highly imbalanced meteorological datasets. The powerful synergy of local spatial gradient extraction (CNN) and sequential dependency mapping (LSTM) perfectly aligned with the physical nature of vertical wind shear, securing an exceptional PR-AUC score of 0.9672, far surpassing the baseline performance.

Finally, through the application of threshold optimization on the Precision-Recall curve, the CNN-LSTM network successfully minimized hazardous missed detections to merely 7 critical cases out of the entire test distribution, while simultaneously suppressing false alarms. This proves its practical viability and technical robustness as a highly dependable computational core for automated flight safety warning systems. Ultimately, this research confirms that while Transformers excel in massive natural language processing tasks, the specific spatial-temporal inductive biases inherent in CNN-LSTM networks make them significantly more suited to decoding the non-linear physics of short atmospheric profiles for extreme weather detection.

ACKNOWLEDGEMENT

The authors express their gratitude to the Meteorology, Climatology, and Geophysical Agency (BMKG) Soekarno-Hatta for providing the historical radiosonde dataset, and the Master of Informatics Engineering Program at Universitas Pamulang for the academic support during this research.

REFERENCES

- [1] International Civil Aviation Organization. (2018). Manual on low-level wind shear and turbulence (Doc 9817, Amendment No. 3). ICAO.

- [2] Ernani, Y., Sukojo, B. M., Ratnasari, Y., & Anggoro, M. D. (2025). Analysis of low-level wind shear (LLWS) triggers using radar and LIDAR data at Soekarno-Hatta International Airport. *IOP Conference Series: Earth and Environmental Science*, 1551(1), 012032.
- [3] Lee, Y.-G., Ryoo, S.-B., Han, K., Choi, H.-W., & Kim, C. (2020). Inter-comparison of ensemble forecasts for low level wind shear against local analyses data over Jeju area. *Atmosphere*, 11(2), 198.
- [4] Zamreeg, A. O., & Hasanean, H. M. (2024). Windshear analysis over six airports in Saudi Arabia. *Discover Sustainability*, 5, 233.
- [5] Chen, J., Xie, C., Ji, J., & Lu, J. (2025). Multi-scale wind shear at a plateau airport: Insights from lidar and radiosonde observations. *Remote Sensing*, 17(16), 2762.
- [6] Zhao, S., Yang, C., Shan, Y., & Zhu, F. (2025). Identification and analysis of wind shear within the transiting frontal system at Xining International Airport using lidar. *Remote Sensing*, 17(5), 732.
- [7] Alves, D., Mendonça, F., Mostafa, S. S., & Morgado-Dias, F. (2024). Low tropospheric wind forecasts in aviation: The potential of deep learning for terminal aerodrome forecast bulletins. *Pure and Applied Geophysics*, 181, 2265–2276.
- [8] Khattak, A., Zhang, J., Chan, P.-w., & Chen, F. (2024). A new frontier in wind shear intensity forecasting: Stacked temporal convolutional networks and tree-based models' framework. *Atmosphere*, 15(11), 1369.
- [9] Khattak, A., Chan, P.-w., Chen, F., & Almaliki, A. H. (2025). Deep ResNet strategy for the classification of wind shear intensity near airport runway. *Computer Modeling in Engineering & Sciences*, 142(2), 1565–1586.
- [10] Ripesi, P., & Criscuolo, P. (2024). Low-level wind shear prediction based on machine learning techniques: A case study of Palermo-Punta Raisi International Airport. *SESAR Innovation Days 2024*, Rome, Italy.
- [11] Hallgren, C., Aird, J. A., Ivanell, S., Körnich, H., Vakkari, V., Barthelmie, R. J., Pryor, S. C., & Sahlée, E. (2024). Machine learning methods to improve spatial predictions of coastal wind speed profiles and low-level jets using single-level ERA5 data. *Wind Energy Science*, 9, 821–840.
- [12] Khattak, A., Zhang, J., Chan, P.-w., & Chen, F. (2024). Assessment of wind shear severity in airport runway vicinity using interpretable TabNet approach and Doppler LiDAR data. *Applied Artificial Intelligence*, 38(1), e2302227.
- [13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- [14] Peng, K., Xin, J., Zhu, X., Cao, X., Wang, Z., Ma, Y., Zhao, D., & Ren, X. (2025). A novel 3D physics-integrated swin-transformer model for precise high-resolution urban boundary layer wind speed estimation. *Geophysical Research Letters*, 52, e2025GL115246.
- [15] Ryan, M., Saputro, A. H., & Sopaheluwakan, A. (2022). Intelligent low-level wind shear alert prediction system based on anemometer sensor network and temporal convolutional network (TCN). *Geographia Technica*, 17(1), 92–103.
- [16] Khattak, A., Chan, P.-w., Chen, F., & Almaliki, A. H. (2025). A hybrid VMD-TPE-TCN framework for wind shear prediction near an airport runway. *Atmosphere*, 16(4), 381.
- [17] Seok, J.-H., Choi, H.-W., & Lee, S.-S. (2026). Wind shear prediction at Jeju International Airport using a tree-based machine learning algorithm. *Forecasting*, 8(1), 4.
- [18] Choi, H.-W., Kim, Y.-H., Han, K., & Kim, C. (2021). Probabilistic forecast of low level wind shear of Gimpo, Gimhae, Incheon and Jeju International Airports using ensemble model output statistics. *Atmosphere*, 12(12), 1643.

- [19] Huang, J., Ng, M. K. P., & Chan, P. W. (2021). Wind shear prediction from light detection and ranging data using machine learning methods. *Atmosphere*, 12(5), 644.
- [20] Xia, H., Chen, Y., Yuan, J., Su, L., Yuan, Z., Huang, S., & Zhao, D. (2024). Windshear detection in rain using a 30 km radius coherent Doppler wind lidar at mega airport in plateau. *Remote Sensing*, 16(5), 924.