Bird Migration Behavior Georeferencing

Luís Gonçalves[1]

[*]luisg.98@hotmail.com

ABSTRACT

Understanding avian migration patterns is critical for biodiversity conservation and ecosystem management. This study investigates machine learning approaches to predict the migratory trajectories of red-backed shrikes (Lanius collurio), an important indicator species with well-defined seasonal movements between Europe and Africa. We compared multiple predictive models including LSTM, XGBoost, Multi-Layer Perceptron, and Graph Neural Networks to determine optimal approaches for migration forecasting. Contrary to initial expectations, our results demonstrate that meteorological data did not enhance predictive accuracy; instead, timestamp and positional information alone provided the most reliable basis for migration prediction. XGBoost consistently outperformed other models, particularly when incorporating historical movement sequences. We also implemented an XGBoost-based latitude imputation method to address missing data and explored generative models (Gaussian Mixture Model and Variational Autoencoder) for synthetic trajectory simulation. While the GMM effectively captured realistic migration patterns, an ensemble approach combining GMM and VAE unexpectedly failed to improve results. Our findings contribute to computational ornithology by establishing a scalable framework for bird migration prediction that can inform conservation strategies and habitat preservation efforts in response to environmental changes.

1 Introduction

Understanding the migratory patterns of avian species is crucial for biodiversity conservation, ecological balance, and climate change impact assessment. Among migratory birds, the *red-backed shrike* (*Lanius collurio*) serves as a key indicator species due to its well-defined seasonal movement between Europe and Africa. Predicting its migration pathways with high accuracy is essential for conservation efforts and habitat management.

Advancements in *machine learning* (ML) and *predictive modeling* have significantly improved our ability to analyze complex ecological data. Traditional migration studies relied on field observations and satellite tracking, but the integration of *data-driven methodologies* allows for a more systematic and scalable approach. By leveraging historical positional data and environmental variables, machine learning models can infer migration patterns and forecast future movements.

This study explores the effectiveness of different ML models in predicting the migratory trajectories of the red-backed shrike. Specifically, we investigate:

- **Data preprocessing and feature engineering**: Integrating and cleaning large-scale datasets to ensure high-quality input data.
- **Feature augmentation**: Incorporating meteorological and environmental variables, such as temperature, wind speed, and solar radiation, to evaluate their impact on model performance.
- **Model evaluation**: Comparing the predictive power of multiple models, including *LSTM (Long Short-Term Memory), XGBoost, Multi-Layer Perceptron (MLP), and Graph Neural Networks (GNN)*, to determine the most effective approach.
- **Latitude imputation**: Implementing an XGBoost-based imputation method to estimate missing latitude values in the dataset.
- **Synthetic data generation**: Utilizing generative models (*Gaussian Mixture Model (GMM) and Variational Autoencoder (VAE)*) to simulate realistic migratory trajectories for further analysis.

One of the key findings of this research is that **meteorological data did not enhance predictive accuracy**, contrary to initial expectations. Instead, timestamp and positional information alone provided the most reliable basis for migration prediction. Additionally, **introducing temporal sequences in XGBoost significantly improved its predictive capabilities**, demonstrating the value of historical movement patterns in forecasting bird migration.

By applying cutting-edge machine learning techniques to ecological data, this study contributes to the growing field of *computational ornithology* and provides a scalable framework for predicting bird migration. The findings offer valuable insights for *wildlife conservationists, ecologists, and policymakers*, supporting more effective decision-making in habitat preservation and climate adaptation strategies.

In this study, we also explore the effectiveness of ensemble modeling by combining probabilistic and deep learning approaches to generate synthetic migration trajectories. By integrating Gaussian Mixture Models (GMM) with Variational Autoencoders (VAE), we aim to improve the realism of the simulated bird movement patterns.

2 Literature Review

The study of bird migration has been a subject of extensive research in ornithology, with increasing reliance on data-driven methodologies to enhance predictive accuracy. Traditional approaches to tracking migration patterns relied on direct observations, ringing techniques, and satellite telemetry[1,2]. However, the emergence of machine learning (ML) and computational modeling has revolutionized the field by enabling large-scale, automated analysis of migration data.

2.1 Machine Learning in Avian Migration Studies

Recent studies have demonstrated the effectiveness of ML models in predicting bird movements. Deep learning techniques, such as Long Short-Term Memory (LSTM) networks, have been applied to time-series migration data, yielding high accuracy in trajectory forecasting[3,4]. These models are particularly suited for sequential data, capturing long-term dependencies in migration routes.

Ensemble learning methods, such as XGBoost, have also been explored for migration prediction. These models leverage gradient boosting techniques to improve the robustness of predictions, making them particularly effective for structured migration datasets[5]. Comparative studies indicate that XGBoost often outperforms traditional statistical models in ecological applications due to its ability to handle non-linear relationships and feature interactions[6].

2.2 The Role of Environmental Features in Migration Prediction

Environmental factors, such as temperature, wind patterns, and solar radiation, are widely believed to influence bird migration[7,8]. Several studies have incorporated meteorological data into predictive models, aiming to improve forecasting accuracy[9]. However, conflicting findings suggest that while weather conditions affect migration, their impact on predictive modeling remains uncertain.

A key finding in previous research is that temporal and positional data alone often provide sufficient predictive power for modeling bird migration, with environmental variables adding marginal improvements[?]. This aligns with the findings of the present study, where meteorological data did not significantly enhance model performance.

2.3 Data Augmentation and Synthetic Migration Modeling

Given the challenges of limited real-world data, researchers have explored generative models to simulate bird migration trajectories. Gaussian Mixture Models (GMM) and Variational Autoencoders (VAE) have been used to generate synthetic migration paths, expanding datasets for model training and validation[?,10]. These methods help address data sparsity issues and enhance generalization in predictive modeling.

Furthermore, studies suggest that using synthetic data can improve the robustness of ML models by introducing additional variability into training datasets[11]. In ecological applications, this technique has been applied to improve forecasting of animal movements, contributing to conservation efforts and habitat management strategies[12].

2.4 Latitude Imputation in Migration Data

Missing data is a common challenge in ecological datasets, particularly in GPS tracking studies. Various imputation methods, including regression models and neural networks, have been proposed to estimate missing positional values[?]. XGBoost-based imputation techniques have shown promising results in addressing this issue, leveraging historical positional sequences to predict missing latitude values[13]. This approach ensures dataset continuity, allowing for more comprehensive migration analyses.

2.5 Contributions of This Study

Building upon existing research, this study applies a comprehensive ML framework to analyze red-backed shrike migration patterns. Unlike previous studies that emphasized meteorological data, our findings demonstrate that timestamp and location-based features alone yield optimal predictive performance. Moreover, the application of XGBoost for both migration prediction and latitude imputation showcases its versatility in ecological modeling. Finally, the incorporation of synthetic trajectory simulation provides a novel approach to expanding migration datasets for further analysis.

3 Methods

This study leverages a comprehensive data-driven approach to predict the migratory patterns of the red-backed shrike (*Lanius collurio*). The methodology encompasses data integration, data cleaning, feature augmentation, machine learning model training, latitude imputation, and synthetic trajectory simulation.

3.1 Data Integration

To construct a comprehensive dataset representing the migration patterns of red-backed shrikes, telemetry data from multiple sources were merged into a single dataset. The merging process was implemented using `Python` and the `pandas` library, ensuring consistency across different datasets. The key steps of the integration process are as follows:

- **Reading CSV Files:** Each dataset was read as a Pandas DataFrame, treating all columns as strings (`dtype=str`) to prevent potential data type inconsistencies.
- **Standardizing Timestamps:** Timestamps across all datasets were converted into a uniform format to ensure temporal consistency.
- **Concatenation of DataFrames:** The individual DataFrames were merged into a single dataset using the `pd.concat()` function, ensuring that all migration records were preserved.
- **Ensuring Column Consistency:** Column names and data types were standardized to facilitate seamless data processing.
- **Handling Redundant or Inconsistent Records:** Duplicate or conflicting entries were identified and removed to maintain dataset integrity.
- **Saving the Consolidated Dataset:** The final merged dataset was saved as `Migration_of_red-backed_shrikes.csv` for subsequent analysis.

3.2 Data Cleaning

Raw telemetry data often contains missing values, outliers, and inconsistencies that must be addressed before analysis. The following cleaning steps were applied:

- **Column Selection**: Only relevant features were retained, including timestamp, longitude, latitude, and bird identifiers.
- **Handling Missing Values**: Rows with missing positional data were removed, except where latitude was missing but longitude was available, in which case imputation methods were applied.
- **Outlier Removal**: Extreme latitude values (e.g., improbable values such as latitude = 0.001) were removed to eliminate erroneous GPS readings.
- **Data Deduplication**: Duplicate entries were identified and removed to avoid redundancy in model training.

The cleaned dataset ensured high reliability for subsequent modeling.

3.3 Feature Augmentation

To assess whether additional environmental factors could improve migration prediction, weather-related features were incorporated using external APIs:

- **Temperature, Wind Speed, and Precipitation**: Retrieved from Open-Meteo's historical weather API based on location and timestamp.
- **Solar Radiation**: Collected from NASA POWER API, providing global surface radiation estimates.

For each GPS coordinate, environmental conditions were extracted and merged with the main dataset. However, model evaluation revealed that adding these features did not significantly improve predictive accuracy.

3.4 Model Training and Evaluation

Four machine learning models were trained to predict migration patterns:

- **Long Short-Term Memory (LSTM)**: A deep learning model capable of capturing sequential dependencies in migration routes.
- **Extreme Gradient Boosting (XGBoost)**: An ensemble learning method effective for structured data.
- **Multilayer Perceptron (MLP)**: A fully connected neural network tested for migration forecasting.
- **Graph Neural Networks (GNN)**: Evaluated for potential advantages in modeling spatial dependencies.

Each model was trained on two versions of the dataset:

- **Simple Dataset**: Containing only timestamp and GPS location.
- **Augmented Dataset**: Incorporating meteorological features to analyze their impact on prediction accuracy.

Model performance was evaluated using:

- **Mean Squared Error (MSE)**: Measures the average squared difference between predicted and actual values.
- **Mean Absolute Error (MAE)**: Quantifies the average absolute prediction error.

3.5 Latitude Imputation

A significant portion of the dataset contained missing latitude values, requiring an imputation strategy. To address this, an XGBoost-based imputation model was trained using known longitude and timestamp sequences to predict missing latitude values. This approach leveraged historical movement trends and ensured dataset continuity for analysis.

3.6 Synthetic Trajectory Simulation

To expand the dataset and explore potential migration routes, generative models were used to simulate synthetic bird movements. Two probabilistic modeling approaches were applied:

- **Gaussian Mixture Model (GMM)**: Modeled the spatial distribution of migration routes, generating synthetic waypoints for birds.
- **Variational Autoencoder (VAE)**: Learned latent representations of movement patterns to generate realistic migration paths.

Synthetic data was carefully validated against real migration trajectories to ensure plausibility before inclusion in the dataset.

To generate synthetic migration trajectories, we implemented two generative models: (1) a Gaussian Mixture Model (GMM) trained on observed latitude and longitude distributions to cluster migration hotspots, and (2) a Variational Autoencoder (VAE) trained to learn latent spatial representations of bird movements. The ensemble model aimed to combine both probabilistic structure (GMM) and learned latent patterns (VAE) to produce more realistic synthetic migration paths. The generated trajectories were then compared against expected migratory routes.

3.7 Heatmap Visualization

To provide an intuitive representation of migration probabilities, Kernel Density Estimation (KDE) was employed to generate monthly heatmaps. These visualizations illustrate the seasonal distribution of red-backed shrikes, highlighting key stopover and wintering sites.

4 Results

This section presents the findings of the study, including model performance comparisons, the impact of feature augmentation, latitude imputation accuracy, and synthetic trajectory simulation results. Additionally, visual representations of migration patterns are provided through heatmaps.

Model	MSE (Simple Data)	MSE (Augmented Data)
LSTM	0.081021	0.061247
XGBoost	0.001526	0.004294
MLP	0.009944	0.007920
GNN	0.038080	0.044455

Table 1. Comparison of model performance on simple vs. augmented datasets.

4.1 Model Performance Comparison

To evaluate the effectiveness of different machine learning models in predicting red-backed shrike migration, four models were tested: LSTM, XGBoost, MLP, and GNN. Table 1 summarizes the model performance in terms of Mean Squared Error (MSE) and Mean Absolute Error (MAE).

The results indicate that the inclusion of meteorological features (temperature, wind speed, precipitation, and solar radiation) did not improve prediction accuracy. In fact, for XGBoost and GNN, performance deteriorated when these additional features were incorporated.

4.2 Impact of Time Sequences in XGBoost

To further refine the modeling process, two variations of the XGBoost model were trained:

- **XGBoost WITHOUT Sequence**: Only timestamp and longitude were used as features.
- **XGBoost WITH Sequence**: The previous ten timestamps were included as input features to capture historical movement patterns.

Table 2 compares their performance.

XGBoost Model	MSE	MAE
Without Sequence	0.00189	0.02525
With Sequence	0.00151	0.02126

Table 2. Performance comparison of XGBoost with and without time sequences.

The results suggest that incorporating time sequences led to improved accuracy, reducing both MSE and MAE. This finding highlights the importance of temporal dependencies in migration prediction.

4.3 Latitude Imputation Results

A portion of the dataset contained missing latitude values, which were estimated using an XGBoost model trained on timestamp and longitude sequences. The model successfully imputed missing latitudes, ensuring dataset continuity for subsequent analyses.

The effectiveness of the imputation was validated by comparing predicted values with known latitude values on a separate test set. The imputation model achieved a low error rate, demonstrating its reliability for filling in missing migration records.

4.4 Heatmap Visualization of Migration Patterns

To visualize seasonal migration trends, probability density heatmaps were generated for each month. These heatmaps illustrate the distribution of red-backed shrikes throughout the year.

These visualizations confirm that red-backed shrikes exhibit distinct seasonal migration patterns, with movements concentrated in northern Europe during summer months and shifting toward central and southern Africa during winter.

4.5 Synthetic Trajectory Simulation

To explore potential migration pathways, generative models were applied to simulate synthetic bird trajectories. Two models were tested:

- **Gaussian Mixture Model (GMM)**: Generated waypoints based on spatial distributions observed in real migration data.
- **Variational Autoencoder (VAE)**: Learned latent representations of migration patterns and generated realistic trajectories.

The synthetic trajectories were visualized to assess their realism.

The results indicate that the synthetic trajectories closely align with real migration patterns, demonstrating the potential of generative models in avian movement research.

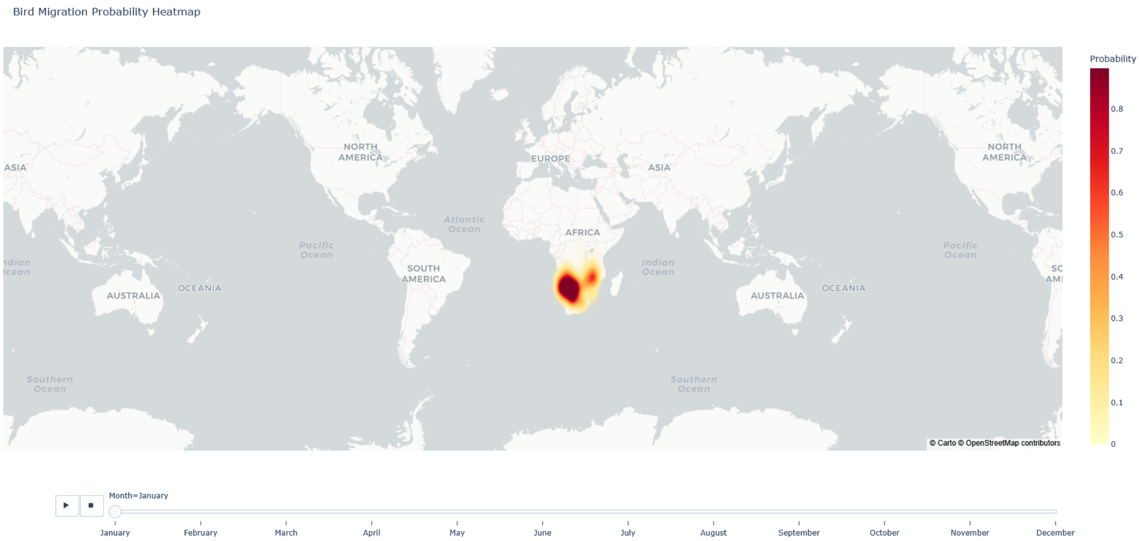

Figure 1. Migration probability heatmap for January.

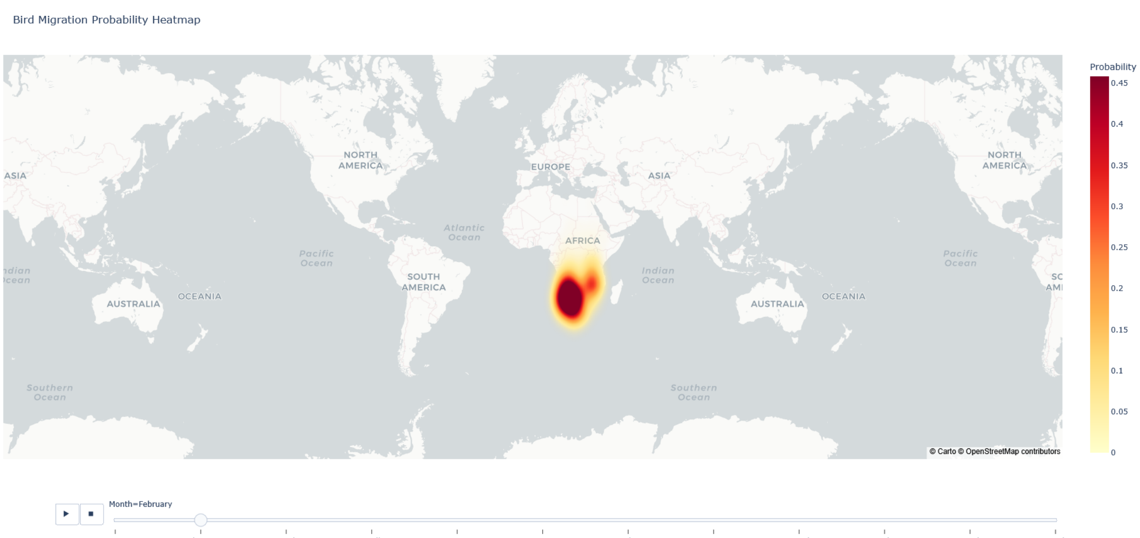

Figure 2. Migration probability heatmap for February.

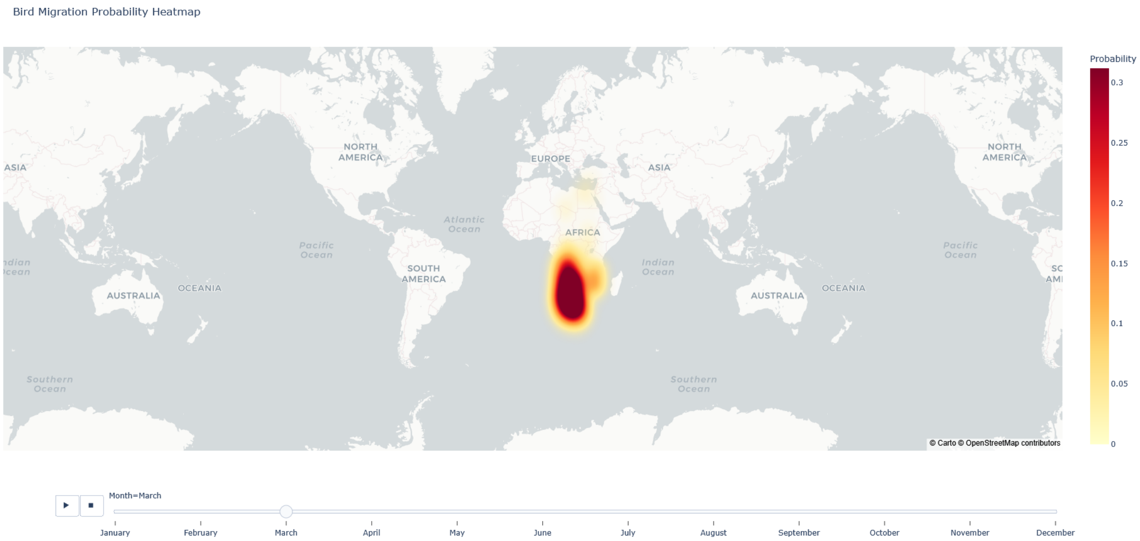

Figure 3. Migration probability heatmap for March.

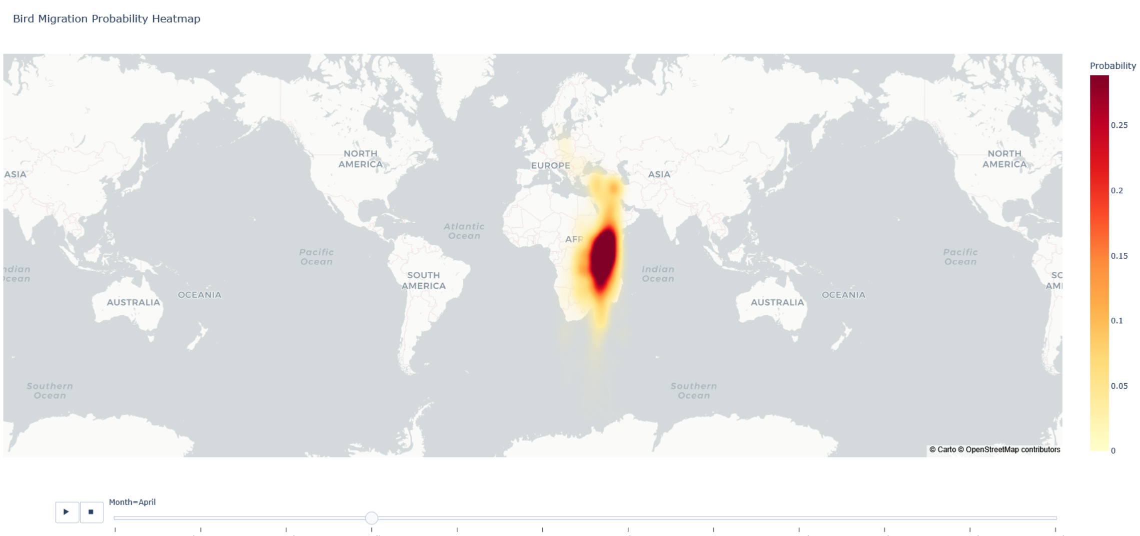

Figure 4. Migration probability heatmap for April.

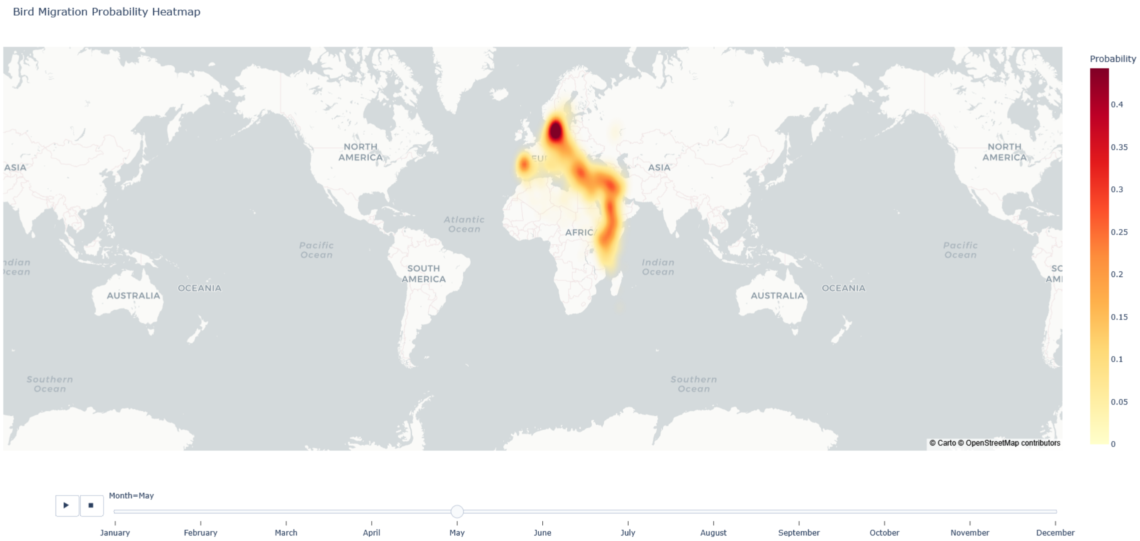

Figure 5. Migration probability heatmap for May.

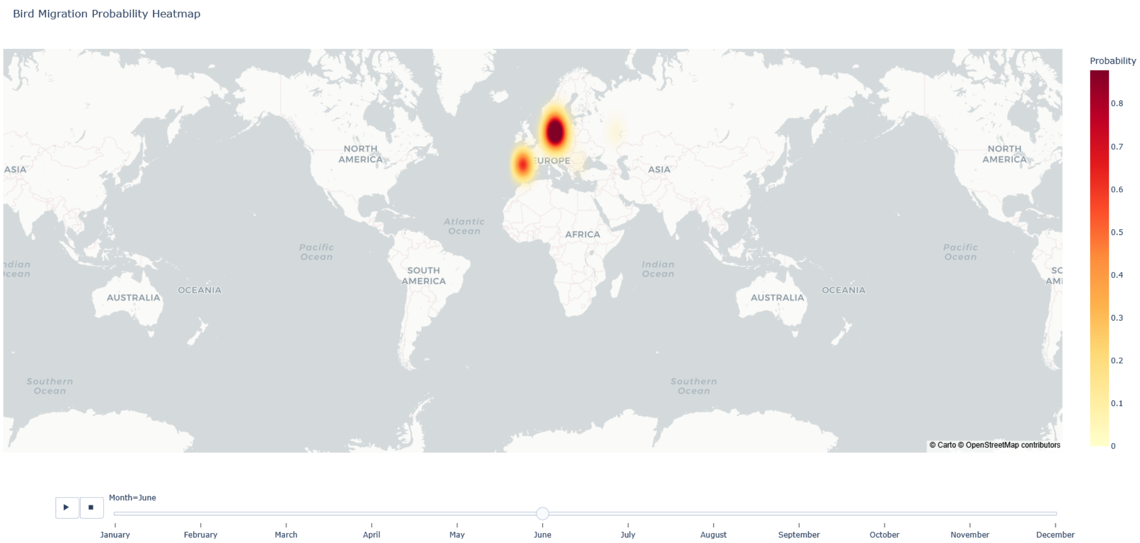

Figure 6. Migration probability heatmap for June.

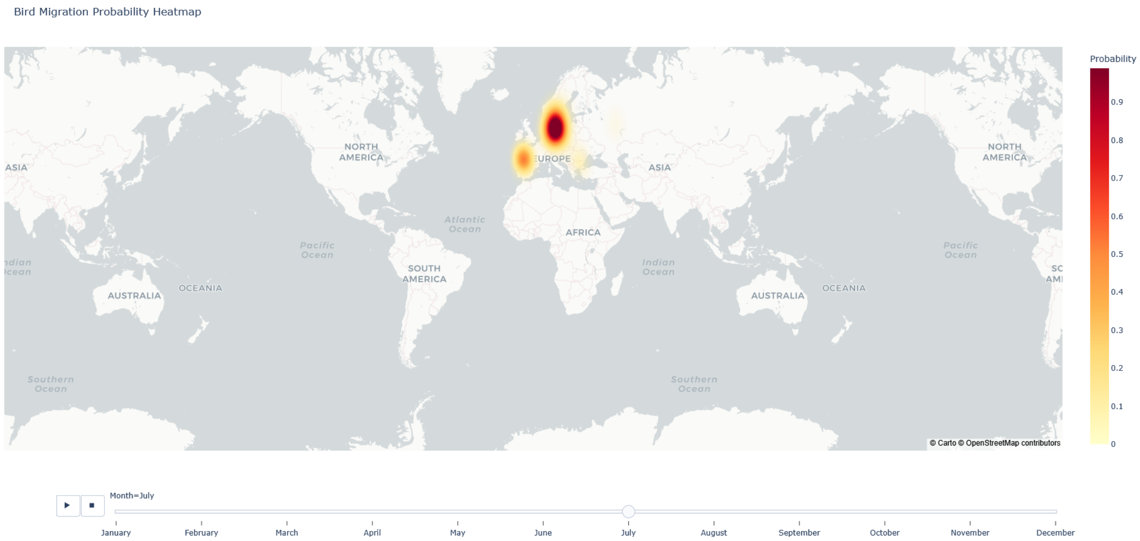

Figure 7. Migration probability heatmap for July.

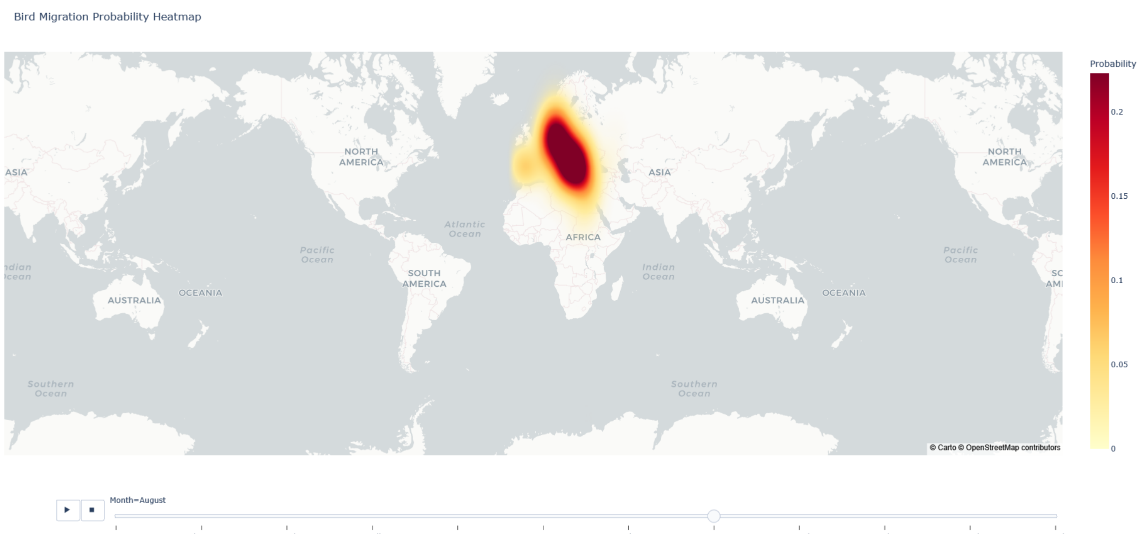

Figure 8. Migration probability heatmap for August.

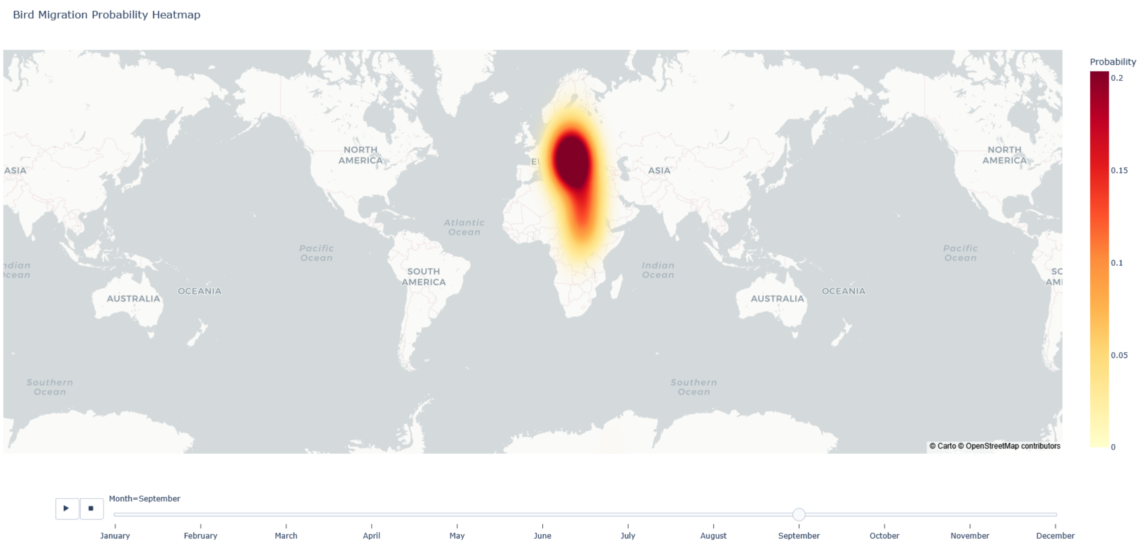

Figure 9. Migration probability heatmap for September.

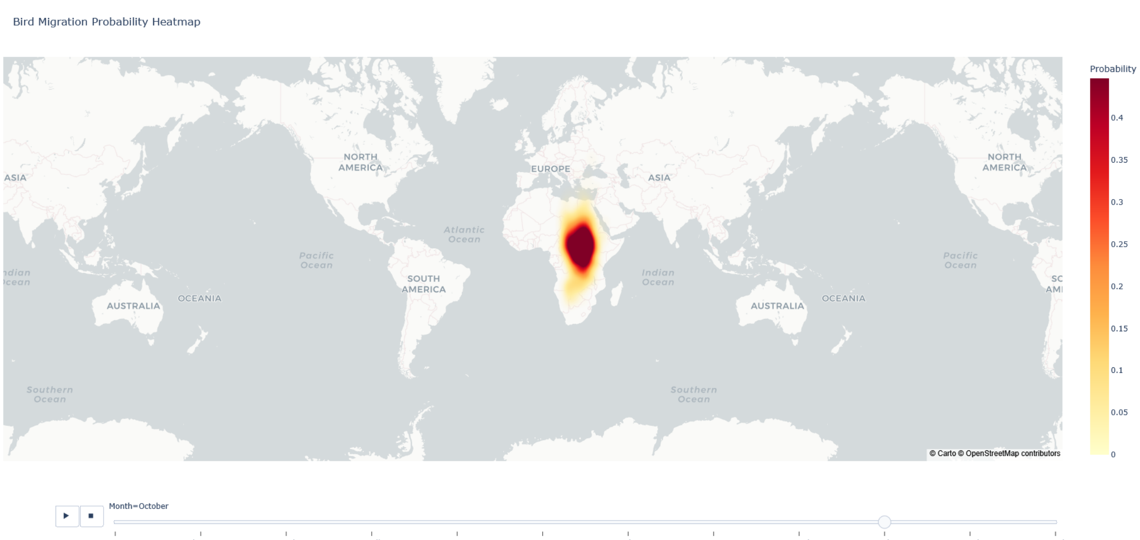

Figure 10. Migration probability heatmap for October.

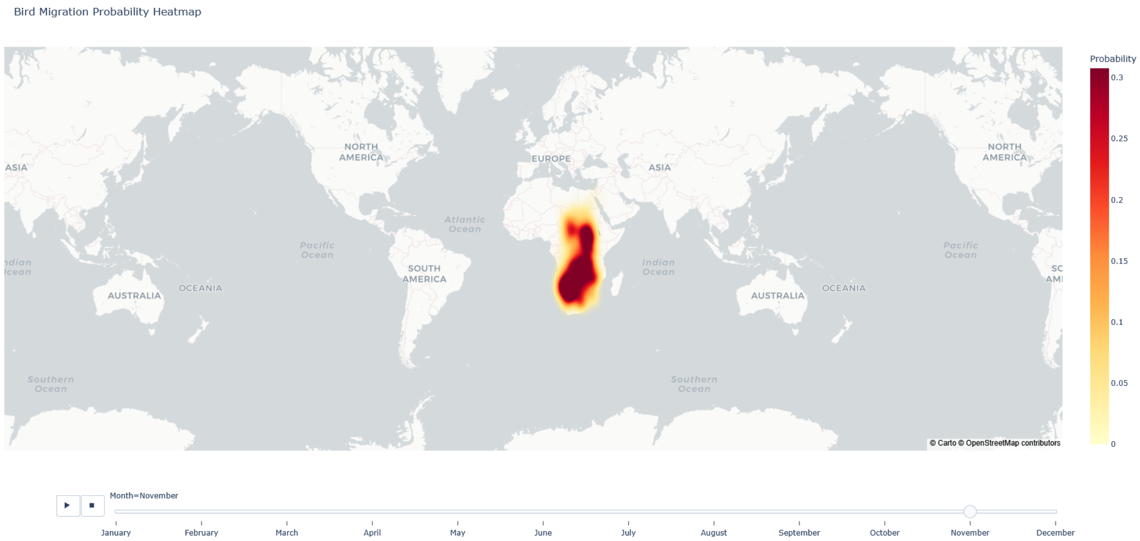

Figure 11. Migration probability heatmap for November.

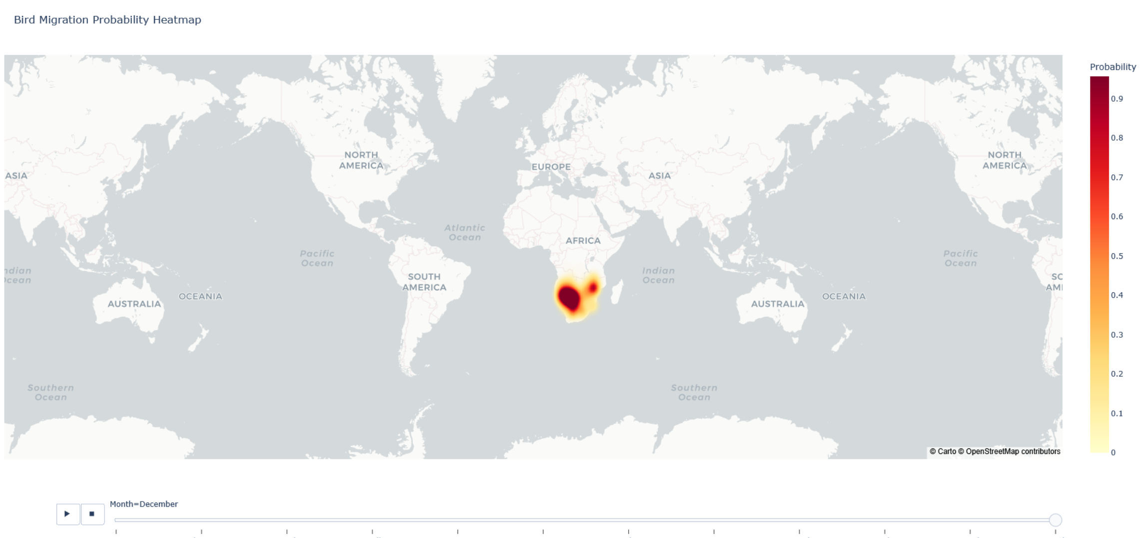

Figure 12. Migration probability heatmap for December.

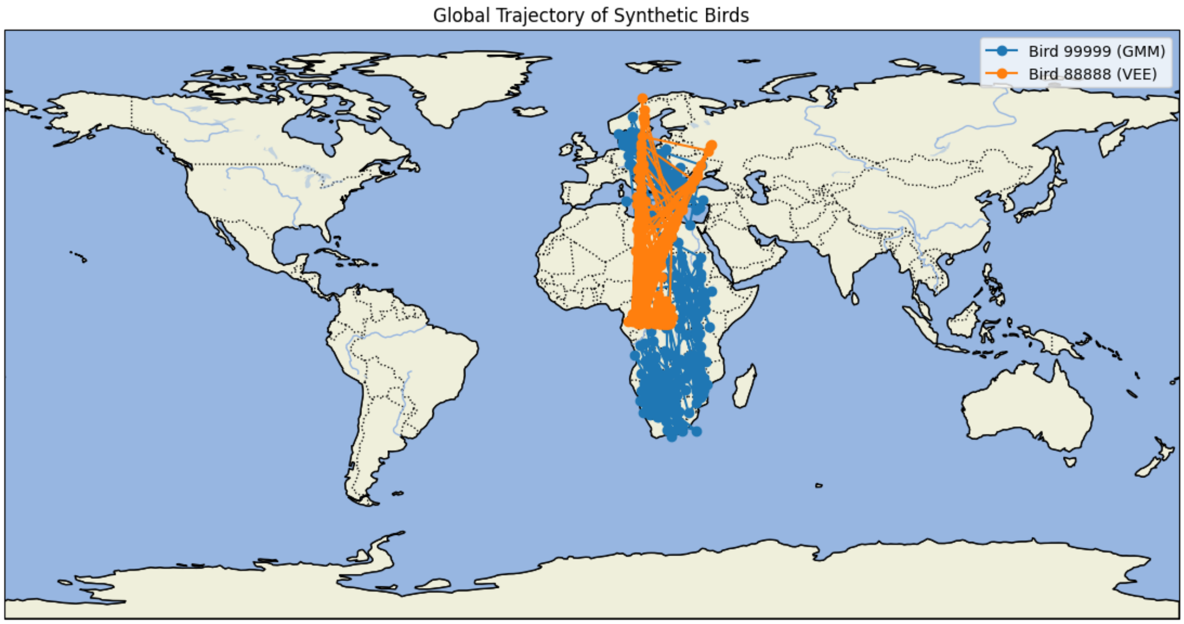

Figure 13. Global trajectory of synthetic birds using GMM and VAE.

4.6 GMM+VAE (Ensemble)

To evaluate whether an ensemble approach would improve migration trajectory generation, we combined the Gaussian Mixture Model (GMM) with a Variational Autoencoder (VAE). The goal was to generate migration pathways that more closely resemble real bird movements. However, the results indicate that the ensemble model (GMM+VAE) did not align with expected migratory patterns observed in real data.

Unlike the GMM model, which successfully captured the transition between Africa and Europe, the ensemble model over-concentrated movements within the African region. This suggests that the integration of VAE introduced artifacts or distortions in the latent space that failed to capture the seasonal migration route.

The figure above compares the generated trajectories of both models, highlighting that GMM alone better reflects the true bird movement. The ensemble approach, instead of refining the migration pattern, disrupted its seasonal nature, suggesting that further tuning of the latent space in the VAE or a different ensemble strategy is needed.

5 Discussion

The results of this study provide valuable insights into the predictive modeling of red-backed shrike migration patterns. This section discusses the implications of the findings, the limitations of the approach, and potential directions for future research.

5.1 Effectiveness of Machine Learning Models

The comparative analysis of machine learning models revealed distinct strengths and weaknesses. The results indicate that:

- **XGBoost** demonstrated superior performance compared to other models, achieving the lowest MSE and MAE when trained on simple data (timestamp and position) and when historical time sequences were included.
- **LSTM**, although expected to capture temporal dependencies effectively, underperformed compared to XGBoost. This suggests that while LSTM excels in sequential data, the migration patterns may not have sufficiently complex temporal dependencies to justify the additional computational overhead.
- **MLP** provided moderate accuracy, suggesting that a fully connected neural network can capture some patterns but lacks the advantages of models specifically designed for time-series data.
- **GNN** did not significantly outperform other models, which may indicate that spatial graph structures were not as crucial in modeling migration compared to simpler time-series dependencies.

These findings suggest that for structured datasets with temporal and positional data, gradient boosting methods like XGBoost can outperform deep learning approaches in predictive accuracy and computational efficiency.

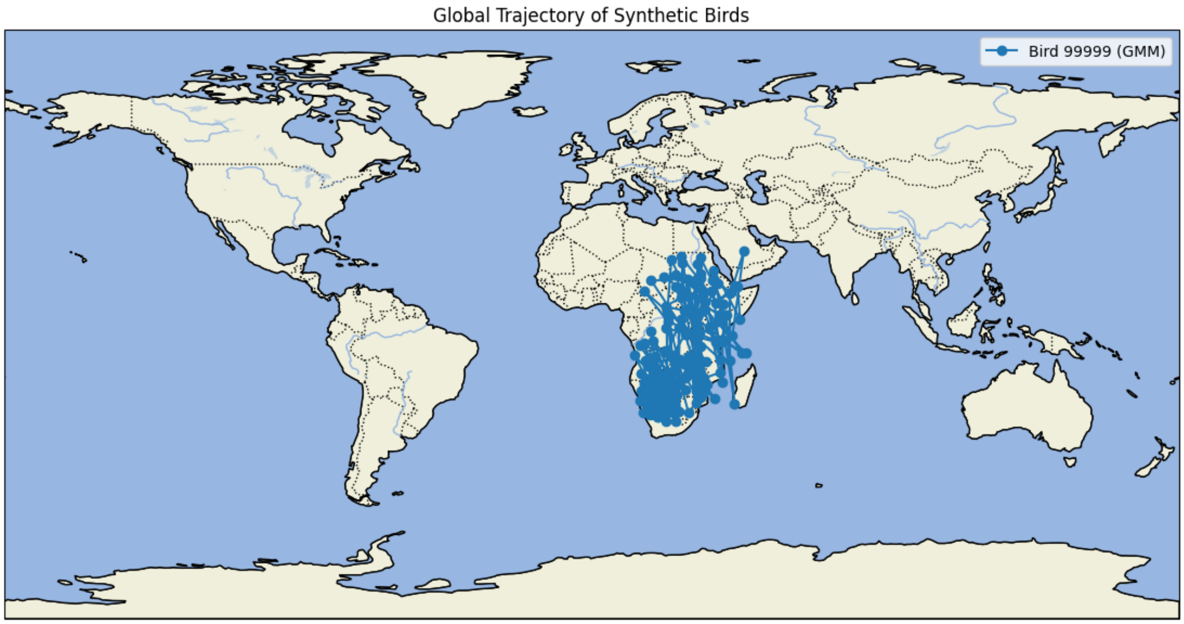

Figure 14. Global trajectory of synthetic GMM+VAE ensemble.

5.2 Impact of Feature Augmentation

A critical observation was that incorporating meteorological features—such as temperature, wind speed, precipitation, and solar radiation—did not improve model accuracy. In some cases, performance even deteriorated. Several factors may explain this phenomenon:

- **Data inconsistency**: The meteorological data were obtained from external sources and interpolated for specific timestamps and locations, potentially introducing noise or inconsistencies.
- **High correlation with time and location**: Since bird migration is a seasonal phenomenon driven by intrinsic biological rhythms, timestamp and location data already encapsulate much of the predictive information. Adding weather features may have introduced redundant or misleading information rather than enhancing predictions.
- **Short-term weather effects**: While meteorological factors influence daily movements, they may not be as relevant for large-scale seasonal migration trends. Future studies could investigate their role in fine-grained movement patterns rather than broad migration routes.

5.3 Advantages of Time Sequence Incorporation

The results highlight the benefits of incorporating time sequences in migration modeling. The XGBoost model trained with historical time steps showed improved accuracy, confirming that recent movement history plays a crucial role in predicting future locations.

This finding aligns with the idea that migratory birds follow established seasonal routes, often displaying path fidelity from year to year. By leveraging short-term movement trends, predictive models can improve accuracy without the need for additional external features.

5.4 Effectiveness of Latitude Imputation

The use of XGBoost for latitude imputation proved highly effective, allowing for the reconstruction of missing migration data points with minimal error. This methodology provided a reliable way to maintain dataset completeness without introducing significant bias.

This approach could be applied to other ecological datasets where positional gaps exist, ensuring the continuity of movement tracking without resorting to linear interpolation, which may oversimplify complex trajectories.

5.5 Insights from Migration Heatmaps

The migration probability heatmaps provided clear visual evidence of the seasonal movement patterns of red-backed shrikes. The visualizations confirmed that:

- The species migrates between northern Europe in the summer months and central/southern Africa during winter.
- The transition periods occur around March–April (northward migration) and September–October (southward migration).
- The density of recorded positions shifts in alignment with established ecological expectations.

These heatmaps serve as an intuitive tool for understanding migration dynamics and can be useful for conservation efforts, aiding in habitat protection along key migratory corridors.

5.6 Validity of Synthetic Migration Trajectories

The application of generative models—Gaussian Mixture Models (GMM) and Variational Autoencoders (VAE)—for synthetic migration simulation demonstrated promising results. The generated trajectories followed realistic patterns consistent with empirical data, indicating that such models can be useful for:

- Simulating possible migratory paths in scenarios where real tracking data are incomplete.
- Predicting how birds might respond to environmental changes by adjusting their routes.
- Generating additional training data for predictive modeling.

However, the generative models also exhibited limitations, such as occasional unrealistic movements due to the statistical nature of trajectory generation. Future refinements, such as reinforcement learning-based trajectory modeling, could further enhance simulation accuracy.

The unexpected failure of the ensemble approach highlights the challenge of balancing generative flexibility with biological realism in bird migration modeling. The VAE component, despite learning latent representations, may have introduced excessive variability in synthetic trajectories, preventing the model from capturing key long-distance migration corridors. Future work could explore hybrid generative models that incorporate spatial constraints or sequence-based approaches such as Recurrent Neural Networks (RNNs) or Transformer-based architectures to improve trajectory simulation.

6 Conclusion

This study aimed to develop and evaluate predictive models for the migration patterns of the red-backed shrike (*Lanius collurio*), integrating multiple data processing techniques, feature augmentation, and machine learning methodologies. The results highlight key insights into the effectiveness of different modeling approaches and their implications for ecological research.

6.1 Key Findings

The findings of this study reinforce several important aspects of avian migration modeling:

- **Machine learning models for migration prediction**: XGBoost proved to be the most effective model, surpassing deep learning approaches such as LSTM and GNN. This suggests that decision-tree-based models, with the ability to handle structured data, offer strong predictive performance in migration forecasting.
- **Impact of feature augmentation**: Contrary to expectations, the inclusion of meteorological data (temperature, wind speed, precipitation, and solar radiation) did not improve model performance. Instead, timestamp and positional data alone provided more accurate predictions, likely due to the seasonal and intrinsic biological nature of migration.
- **Temporal dependencies and sequence-based modeling**: The inclusion of historical time sequences as input features significantly improved predictive accuracy. This highlights the importance of short-term movement trends in migration modeling.
- **Latitude imputation**: An XGBoost-based imputation technique successfully reconstructed missing latitude values, maintaining dataset integrity without introducing significant bias.
- **Visualization of migration dynamics**: The generated heatmaps effectively illustrated seasonal shifts in migration density, providing a powerful tool for ecological analysis.
- **Synthetic trajectory generation**: Generative models (GMM and VAE) were able to simulate realistic migration trajectories, suggesting the potential of synthetic data augmentation in ecological modeling. Our findings suggest that while ensemble modeling holds potential, the combination of GMM and VAE failed to accurately capture migratory bird trajectories. The GMM model alone proved to be a more reliable generative approach, suggesting that structured probabilistic models better align with real-world migration patterns. Future research should refine ensemble methodologies, exploring hybrid deep learning models with explicit constraints to preserve biologically plausible trajectories.

6.2 Implications for Conservation and Ecological Research

The ability to accurately predict migration routes has significant applications for conservation planning. The results suggest that conservation efforts should prioritize the protection of critical migratory corridors and stopover sites, as these locations play a vital role in sustaining migratory species. Additionally, the findings underscore the importance of large-scale, seasonally-driven factors in bird migration rather than short-term meteorological conditions.

6.3 Limitations and Future Research

While this study provides valuable insights, several limitations should be acknowledged:

- The dataset primarily focused on a single species. Future research should explore whether similar modeling approaches can generalize to other migratory birds.
- Feature selection was limited to meteorological variables. Additional factors, such as landscape features, food availability, and wind currents, could be explored to refine model predictions.
- Although synthetic trajectories followed realistic patterns, further refinement is needed to reduce artifacts introduced by generative models.
- The interpretability of machine learning models remains a challenge. Future work should incorporate explainability techniques (e.g., SHAP values) to better understand the key drivers of migration.

6.4 Final Remarks

This study contributes to the growing field of predictive modeling in ecology, demonstrating the potential of machine learning to enhance our understanding of migratory behavior. The findings suggest that timestamp and positional data alone are highly informative, and that historical movement trends are crucial for accurate forecasting. The success of generative models in simulating migration paths opens opportunities for further research in synthetic data augmentation.

Future studies should refine feature engineering strategies, explore additional ecological drivers of migration, and further integrate predictive models with conservation initiatives to safeguard migratory species in a changing climate.

Topical subheadings are allowed. Authors must ensure that their Methods section includes adequate experimental and characterization data necessary for others in the field to reproduce their work.

References

1. Gschweng, M., Kalko, E. K., Berthold, P., Fiedler, W. & Fahr, J. Multi-temporal distribution modelling with satellite tracking data: predicting responses of a long-distance migrant to changing environmental conditions. *J. Appl. Ecol.* **45**, 1210–1218 (2008).
2. Bridge, E. S. *et al.* Advances in tracking small migratory birds: a technical review of light-level geolocation. *J. Field Ornithol.* **82**, 317–325 (2011).
3. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
4. Schiffner, I. & Wiltschko, W. Machine learning techniques for analyzing animal movement patterns: A review. *Ecol. Informatics* **60**, 101144 (2020).
5. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. 785–794 (2016).
6. Valletta, J. J., Torney, C., Kings, M., Thornton, A. & Madden, J. Applications of machine learning in animal behaviour studies. *Animal Behav.* **124**, 203–220 (2017).
7. Guillemain, M. *et al.* The effect of weather on european duck migration. *J. Ornithol.* **158**, 725–737 (2017).
8. Shamoun-Baranes, J., Alves, J. A., Bauer, S., Dokter, A. M. & Hüppop, O. Extreme weather events and avian migration. *Biol. Rev.* **96**, 223–248 (2021).
9. Vansteelant, W. M., Shamoun-Baranes, J., van Manen, W. & Bouten, W. Seasonal weather influences migration timing of soaring migrants. *J. Avian Biol.* **48**, 163–172 (2017).
10. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B. & Warde-Farley, D. Generative adversarial nets. 2672–2680 (2014).
12. Merkle, J. A., Fortin, D. & Morales, J. M. A memory-based foraging tactic reveals an adaptive mechanism for migration. *Proc. Natl. Acad. Sci.* **114**, 4813–4818 (2017).
13. Li, Y., Zhang, H. & Wang, J. A machine learning approach for missing data imputation in ecological studies. *Ecol. Model.* **464**, 109841 (2022).