BAYESIAN MODELLING OF MULTI-YEAR CROP TYPE CLASSIFICATION USING DEEP NEURAL NETWORKS AND HIDDEN MARKOV MODELS

Gianmarco Perantoni, Giulio Weikmann, Lorenzo Bruzzone

Department of Information Engineering and Computer Science, University of Trento, Italy

ABSTRACT

The temporal consistency of yearly land-cover maps is of great importance to model the evolution and change of the land cover over the years. In this paper, we focus the attention on a novel approach to classification of yearly satellite image time series (SITS) that combines deep learning with Bayesian modelling, using Hidden Markov Models (HMMs) integrated with Transformer Encoder (TE) based DNNs. The proposed approach aims to capture both i) intricate temporal correlations in yearly SITS and ii) specific patterns in multiyear crop type sequences. It leverages the cascade classification of an HMM layer built on top of the TE, discerning consistent yearly crop-type sequences. Validation on a multiyear crop type classification dataset spanning 47 crop types and six years of Sentinel-2 acquisitions demonstrates the importance of modelling temporal consistency in the predicted labels. HMMs enhance the overall performance and F1 scores, emphasising the effectiveness of the proposed approach.

***Index Terms*—** deep learning, multitemporal classification, HMM, satellite image time series, remote sensing

1. INTRODUCTION

In the last decade, Deep Learning (DL) has become more and more relevant in remote sensing data analysis, allowing the research community to move towards more complex challenges on larger scales. However, with single acquisitions, it is not possible to model many characteristics of the land covers that are strictly related to their temporal signatures, such as in the case of cultivation or blooming vegetation. When multitemporal acquisitions are available, the use of DL allows the modelling of the phenological traits of the land cover [1]. The DL architectures considered for multitemporal remote sensing usually discard the spatial information and focus on the spectral and temporal information. They can be divided into three categories: i) Recurrent Neural Networks (RNNs), ii) 1D Convolutional Neural Networks (CNNs) and iii) Transformer networks. Among the RNNs, the Long-Short-Term Memory (LSTM) network showed high accuracies [2]. In [3], the authors proposed a Temporal CNN (TempCNN) that applies 1D convolutions to the temporal domain, outperforming methods based on random forests and RNNs with Gated Recurring Units (GRUs). Transformer, which are based on the concept of Self-Attention, showed excellent performance in the natural language processing domain, and they are currently the most studied type of Deep Neural Networks (DNNs) for satellite image time series (SITS) analysis. Rußwurm *et al.* [4] compared Transformers, LSTM networks and TempCNNs on a crop type classification task with Sentinel-2 (S2) multispectral time series (TS). They showed that all these models perform similarly on pre-processed data (*i.e.*, with atmospheric correction, temporal selection of cloud-free observations, and cloud masking), whereas LSTM networks and Transformers perform better on raw Top-of-Atmosphere (TOA) data.

A typical application of multitemporal acquisitions is the analysis of SITS over the years to produce yearly land-use/land-cover (LULC) maps. This application requires that multitemporal LULC products outline a realistic temporal evolution of each pixel. For this reason, the temporal consistency of multitemporal land-cover products is particularly important. A Bayesian approach to the classification of two images acquired over the same location was proposed by Swain in his pioneering work [5]. He specifically modelled the temporal correlation between the land cover at different acquisition times to relate the class posterior probabilities estimated on single acquisitions. Then, he proposed a cascade implementation that, starting from a single time step, uses the temporal correlation modelled by known land-cover transition probabilities to account for temporal consistency with the second time step. Bruzzone [6] extended this work to a compound classification system, proposing automatic techniques for the estimation of the transition probabilities and generalising the system to multisensor data and multiple classification systems (e.g., KNNs, SVMs and shallow neural networks) [7]. Extensions to more than two observations usually rely on the Hidden Markov Model (HMM) formulation [8], which allows modelling the temporal correlation through the transition probabilities between adjacent time steps. Nowadays, the classification of multiple dates with multitemporal data and DL-based architectures is still a challenge, and the modelling of the temporal correlation in the TS of the labels predicted by the DNN is understudied in remote sensing. Few works [9, 10, 11, 12, 13] investigated the combination of HMMs with DNNs for acoustic data classification and ECG signal analysis. In remote sensing land cover classification, some papers [14]

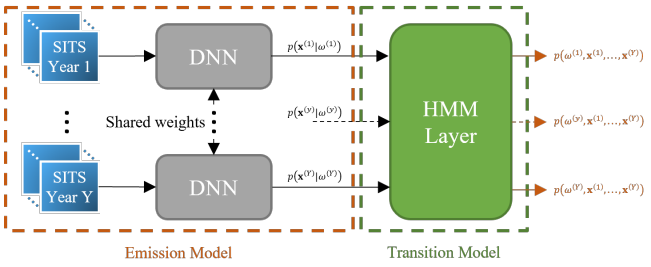

Fig. 1. Flowchart of the proposed cascade-based classification approach, combining a DNN model with an HMM layer.

tried to overcome the complexity of the problem by training an end-to-end DL model, such as an LSTM network, with a multiyear training dataset, defined by the automatic extraction of reliable training samples from existing land cover products. However, the temporal correlation of the labels at different years is not modelled.

In this paper, we focus on the combination of DNNs with strategies for modelling the temporal correlation in the TS of predicted labels, drawing inspiration from a cascade classification approach [5]. The strength of our approach lies in the fusion of Bayesian modelling, specifically HMMs, with DNNs based on the Transformer Encoder (TE). By combining the strengths of these methodologies, we aim to capture both the intricate temporal correlations in multiyear land-cover labels, and the complex spectral patterns within yearly SITS. This strategy is validated on a multiyear crop type classification dataset spanning over six years of S2 acquisitions. The results show the importance of modelling the temporal consistency of the labels predicted by DNNs, as well as the effectiveness of their combination with HMMs.

2. PROPOSED METHODOLOGIES

In this section, we present the novel methodology based on cascade classification employed for multiyear crop type classification, drawing inspiration from HMMs and incorporating a TE as the emission model (Fig. 1). Let $\mathbf{x}^{(y)} \in \mathbb{R}^{T \times B}$ be a TS of T observations with B spectral bands for year y. The task is to predict labels $\omega^{(y)} \in \{\omega_1, \omega_2, ..., \omega_C\}, y = 1, 2, ..., Y$, where C is the number of classes and Y is the number of years under consideration. Under a Bayesian perspective, the problem can be formulated as finding the best sequence of labels that maximise the joint density distribution:

$$p\left(\omega^{(1)}, ..., \omega^{(Y)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(Y)}\right) = p\left(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(Y)} \mid \omega^{(1)}, ..., \omega^{(Y)}\right) P\left(\omega^{(1)}, ..., \omega^{(Y)}\right), \quad (1)$$

where the two terms model the likelihood of the observed sequence given the generating sequence of labels and the prior probability of the sequence of labels, respectively. In the cascade classification approach [5], we consider class-conditional independence in the temporal domain, *i.e.*, the observations $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ at a given spatial position at different time steps y_1, y_2 are independent of each other given the classes $\omega^{(1)}, \omega^{(2)}$ at the two time steps. Then, we can write as follows:

$$p\left(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(Y)} \mid \omega^{(1)}, ..., \omega^{(Y)}\right) = \left[\prod_{y=1}^{Y} p\left(\mathbf{x}^{(y)} \mid \omega^{(y)}\right)\right], \quad (2)$$

which is equivalent to the output independence assumption of HMMs [8]. For two dates classification problems, the cascade classification approach [5] finds the best label $\omega^{(2)}$ by first marginalising $p\left(\omega^{(1)}, \omega^{(2)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right)$ over $\omega^{(1)}$:

$$\arg\max_i \left\{ p\left(\mathbf{x}^{(2)} \mid \omega_i^{(2)}\right) \times \sum_{j=1}^{C} P\left(\omega_i^{(2)} \mid \omega_j^{(1)}\right) p\left(\mathbf{x}^{(1)} \mid \omega_j^{(1)}\right) P\left(\omega_j^{(1)}\right) \right\}. \quad (3)$$

One can note that this is akin to an HMM where the Markov assumption is employed to simplify the definition and modelling of the prior term $P\left(\omega^{(1)}, ..., \omega^{(Y)}\right)$. Combining this with the HMM's output independence assumption and expanding to more than two dates, we perform cascade classification by using a custom implementation of an HMM, where a DNN is used to approximately estimate the emission model $p\left(\mathbf{x} \mid \omega\right)$ and a Markov transition model is used to enforce temporal consistency.

2.1. Emission Model: Transformer Encoder

The proposed approach exploits a TE for raw optical TS classification [4] as the emission model. This DNN is designed to capture complex temporal and spectral patterns present in multitemporal SITS data. Two versions of the TE are considered, differing in the definition of the final layer: the first one uses a standard linear layer with softmax activation, while the second one uses a normalised softmax layer with no bias term and normalised rows in the weight matrix [15]. The former, considered as a discriminative approach, approximately estimates class posterior probabilities $P_{disc}\left(\omega \mid \mathbf{x}\right)$ through standard categorical cross entropy optimisation, while the latter, akin to a generative model, implicitly enforces a uniform class prior distribution $P\left(\omega\right) = 1/C, \forall \omega$ on the class posteriors, aligning with the requirements of an HMM emission model, *i.e.*, $P_{gen}\left(\omega \mid \mathbf{x}\right) \sim p\left(\mathbf{x} \mid \omega\right)$. This approach helps in reducing the bias towards dominant classes in the dataset.

2.2. Transition Model: HMM Layer

To embody the HMM logic, a custom layer is devised to serve as a unique implementation of the HMM transi-

tion model for multiyear crop type classification. The parameters of this HMM layer include the log probabilities $P\left(\omega^{(y)}, \omega^{(y+1)}\right)$ of the joint prior distribution of each tuple of labels $\left(\omega^{(y)}, \omega^{(y+1)}\right)$, representing the probabilities of the two labels co-occurring in consecutive years. This configuration facilitates the extraction of both transition probabilities and initial state probabilities:

$$\begin{aligned} P\left(\omega^{(y+1)} \mid \omega^{(y)}\right) &= P\left(\omega^{(y+1)}, \omega^{(y)}\right) / P\left(\omega^{(y)}\right) \\ P\left(\omega^{(y)}\right) &= \sum_{i=1}^{C} P\left(\omega_i^{(y+1)}, \omega^{(y)}\right), \end{aligned} \tag{4}$$

and improves numerical stability during both training and inference. Moreover, the modelling of the joint probabilities enables the estimation of both forward and backward transition probabilities, simply marginalising over $\omega^{(y)}$ instead of $\omega^{(y+1)}$. Leveraging this capability, the network performs cascade classification in both forward and backward directions during the inference step. This unique approach ensures that the predictions for each year depend on observations from all years, enhancing the overall consistency of predictions. The forward HMM model can be written as follows:

$$\begin{aligned} &p\left(\omega^{(1)}, ..., \omega^{(Y)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(Y)}\right) = \\ &= \left[\prod_{y=1}^{Y} p\left(\mathbf{x}^{(y)} \mid \omega^{(y)}\right)\right] \left[\prod_{y=1}^{Y-1} P\left(\omega^{(y+1)} \mid \omega^{(y)}\right)\right] P\left(\omega^{(1)}\right). \end{aligned} \tag{5}$$

In multiyear cascade classification, the objective is to estimate for each time step i the probability of classes $\omega^{(y)}$ occurring with the observed sequence up to that point:

$$\begin{aligned} &p\left(\omega^{(y)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(y)}\right) = \\ &= \sum_{i_j=1, \forall j<y}^{C} p\left(\omega_{i_1}^{(1)}, \omega_{i_2}^{(2)}, ..., \omega^{(y)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(y)}\right) \\ &= p\left(\mathbf{x}^{(y)} \mid \omega^{(y)}\right) \sum_{i=1} P\left(\omega^{(y)} \mid \omega_i^{(y-1)}\right) \times \\ &\qquad \times p\left(\omega_i^{(y-1)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(y-1)}\right), \end{aligned} \tag{6}$$

where the equation is applied recursively and $p\left(\mathbf{x}^{(1)}, \omega^{(1)}\right) = p\left(\mathbf{x}^{(1)} \mid \omega^{(1)}\right) P\left(\omega^{(1)}\right)$. Assuming that the observations for successive time steps are available, the same approach can be used in the backward direction, obtaining $p\left(\omega^{(y)}, \mathbf{x}^{(y)}, ..., \mathbf{x}^{(Y)}\right)$. The forward and backward predictions can be fused, obtaining a prediction for $\omega^{(y)}$ that depends on the observation of all the years:

$$\begin{aligned} &p\left(\omega^{(y)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(Y)}\right) = \\ &= \frac{p\left(\omega^{(y)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(y)}\right) p\left(\omega^{(y)}, \mathbf{x}^{(y)}, ..., \mathbf{x}^{(Y)}\right)}{p\left(\mathbf{x}^{(y)} \mid \omega^{(y)}\right) P\left(\omega^{(y)}\right)}. \end{aligned} \tag{7}$$

This can be used both for training and inference, where class-posterior probabilities can be obtained simply by normalisation. In this paper, we also explore the second-order HMM, where the transition probabilities are defined over combinations of the two previous time steps, *i.e.*, we have $P\left(\omega^{(y+2)} \mid \omega^{(y+1)}, \omega^{(y)}\right)$. With some adjustments, the cascade equations can be formulated similarly to (6), and the results of forward and backward cascade predictions can be formulated exactly as in (7). Note that, differently from standard HMMs, we do not assume that the same transition model can be used for all the adjacent pairs of years, *i.e.*, we estimate a transition matrix for each couple of years.

2.3. Training Procedure

The emission model is trained encompassing labels from all years. It is important to note that the TE is specifically trained on individual yearly sequences rather than on the entire sequence collectively. The HMM layer is initialised with co-occurrence matrices derived from the TE training phase. The entire model, comprising the TE and HMM layer, undergoes fine-tuning (FT), which is performed in a supervised manner by employing the cascade results of applying the forward-backward cascade algorithm in the categorical cross entropy.

3. DATASET DESCRIPTION

To validate the proposed methodologies, we considered publicly available Land Parcel Identification System (LPIS) crop type maps in Austria, which are based on farmers' declarations[1]. The dataset has been developed with a focus on the Military Grid Reference System (MGRS) tile 33UVP. The selection of the tile was due to the consistent coverage of two S2 orbits and the wide variety of crop types present in the region. We analysed six target agronomic years, ranging from September 2016 to September 2022. The study involved a total number of 805 S2 L1C acquisitions, incorporating all the 13 spectral channels re-sampled at a resolution of $10m$ using a nearest neighbour approach. No cloud screening has been applied to the S2 images. For each agronomic year, we selected the crop fields with a minimum size of $400m^2$ and only the crop fields occurring in every target year with an occurrence greater than 0.05%. At the end of this filtering process, we identified 47 distinct crop type labels, for a total number of $\sim 300,000$ crop fields. Each resulting crop field is then collapsed into one single entry representing the spatial mean value of the pixels belonging to that crop field on every acquisition date.

[1]Data source: `https://www.data.gv.at/en/search/?typeFilter%5B%5D=dataset&searchterm=INVEKOS+Schl%C3%A4ge+%C3%96sterreich+&searchin=data`, Accessed on: May 23, 2024.

Table 1. Comparison of the mean F1 scores (mF1%) considering the different components of the methodology proposed on test set. Results from both training and FT are reported.

	approach	mF1%	FT mF1%
TE	generative	68.71	70.88
	discriminative	64.85	70.05
HMM 1st order	cascade	70.16	73.59
HMM 2nd order	cascade	70.08	73.49

4. EXPERIMENTAL SETUP AND RESULTS

To accurately model the temporal sequences and the consistency of the crop rotations, we split the dataset considering a stratified random sampling approach, considering the uniqueness of the label sequences within the agronomic target years. In particular, we defined each stratum exploiting an agglomerative hierarchical clustering approach, which recursively merges pairs of label sequences based on a Hamming distance with rotations. To determine the distance of two samples, we consider the minimum of each possible hamming distance computed on one sample and all the circular shifts of the other candidate sample. This approach results in the aggregation of samples showing similar crop sequences but shifted by one or more years, as they will display a small distance metric and be grouped within the same cluster. After the clustering operation, each sample in the stratum is randomly assigned to train, validation, and test sets, with a probability of 0.6, 0.2, and 0.2, respectively. The TE was trained on the training set, while the FT of the TE and HMM was conducted considering the performance on the validation set.

Table 1 shows a comparison of the performance of the different models on the test set, before and after FT. HMMs before FT are simply initialised with the co-occurrence matrices obtained from the training labels. From the table, one can see that the TE alone performs worse compared to the combination with an HMM, emphasising the importance of exploring the correlation in the temporal sequence. Moreover, the FT cascaded HMM achieves better performance compared to the inference-only approach (*i.e.*, without FT), showing a 73.59% F1 score when considering the 1st order approach. This is a notable achievement given the complexity of considering 47 different classes—which is a very challenging problem in crop-type mapping with remote sensing image classification. The modified version of the TE employing the normalised weights in the final linear layer (*i.e.*, generative) outperforms the original implementation of the TE (*i.e.*, discriminative) [4]. A more detailed analysis can be inferred from Table 2, where the F1 scores are presented for each distinct crop type, considering the generative TE and both 1^{st} and 2^{nd} order HMMs. Despite the similar performance on average, the classification results indicate that the 2^{nd} order HMM performs better than the 1^{st} order HMM on the most difficult classes at the expenses of the most performing ones.

5. CONCLUSIONS

In this paper, we have presented a novel cascade classification approach to multitemporal land cover classification of SITS that integrates Bayesian modelling, specifically HMMs, with DNNs. This methodology involves two key aspects: (i) the supervised training of an TE used as the emission model of an HMM and (ii) the FT of the cascade classification approach considering an HMM model built on top of the Encoder. This study showcased that the incorporation of HMM with DNNs effectively leverages the consistency of labels across different time sequences, outperforming the independent classification of each single year with the Encoder. The effectiveness of this novel strategy is validated through the experimental results on a multiyear dataset spanning six years of S2 acquisitions on a high detailed crop classification scheme, showing 47 different crop types. Our results highlight the importance of modelling the temporal consistency of labels predicted by DNNs, by fusing it with HMMs in a Bayesian approach, in enhancing the overall performance of crop type classification models. In future works, we plan to reframe the approach within a change detection framework and investigate the weak and semi supervision scenarios, where only a subset of the training data are reliably labelled.

Table 2. Quantitative analysis in terms of F1 scores (%) on the 47 selected crop type classes comparing the fine-tuned generative TE (GTE) and HMMs, considering both 1st (HMM-1) and 2nd (HMM-2) order cascade classification approaches. Best results are highlighted in gray.

Class Name	WINTER BARLEY	PASTURE THREE USES	PASTURE TWO USES	SILAGE MAIZE	WINTER TRITICALE	GREEN FALLOW	SINGLE-MOWN MEADOW	SUMMER OATS	BROAD BEANS	WINTER COMMON WHEAT	GRAIN MAIZE	WINTER MIXED CEREALS	TABLE POTATOES	DRESS	STRAWBERRIES	CLOVER	MISCELLANEOUS FIELD FODDER	FORAGE GRASSES	PERENNIAL PASTURE	ELEPHANT GRASS	GRASSLAND FALLOW	SPRING BARLEY	WINTER RAPESEED	SUMMER MIXED CEREALS	WINTER RYE	SUMMER COMMON WHEAT	SOYBEANS	WINTER SPELT	CORN COB MIX	ALTERNATING MEADOW	RANGELAND	OIL PUMPKIN	ENERGY WOOD	ALPINE FORAGE	TABLE APPLES	PEAS - CEREALS MIXTURE	OTHER ARABLE LAND	GRAIN PEAS	LUCERNE	WINTER CUMIN	SUMMER POPIES	SUNFLOWERS	SINGLE-CULTURE FIELD VEGETABLES	STARCH INDUSTRIAL POTATOES	SEED POTATOES	SUGAR BEETS	SEED MAIZE PROPAGATION
GTE	98.4	86.5	48.3	86.8	88.0	76.7	40.6	83.6	92.9	96.4	87.7	46.2	73.3	62.3	86.4	39.7	22.2	34.0	70.6	97.2	31.3	87.8	99.1	43.5	90.2	58.0	97.8	82.6	28.0	43.4	40.8	92.1	81.5	93.8	82.9	41.5	37.7	71.5	37.7	93.8	83.0	81.1	75.3	86.7	65.9	98.8	87.3
HMM-1	98.3	89.0	56.4	87.9	88.9	83.5	53.3	84.1	92.4	96.3	87.9	44.7	75.5	65.3	91.6	44.1	32.4	33.2	75.6	98.3	44.8	87.6	99.2	43.6	90.3	59.7	97.9	85.2	27.7	48.4	51.0	93.0	87.7	95.1	82.7	38.0	48.3	72.1	41.6	96.2	85.1	77.9	82.5	88.3	68.9	98.9	88.4
HMM-2	98.3	89.0	57.2	89.4	89.5	82.3	52.9	83.8	92.1	96.3	89.4	50.1	73.7	65.6	90.9	44.5	32.7	35.7	76.6	97.1	46.5	87.8	98.8	45.5	90.4	55.6	97.4	83.0	31.4	51.1	48.7	91.5	90.8	95.5	94.1	37.4	38.1	66.2	43.4	92.4	83.8	78.4	80.3	86.1	68.0	98.5	86.2

6. REFERENCES

[1] P. Ghamisi, B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. Chi, K. Anders, R. Gloaguen, P. M. Atkinson, and J. A. Benediktsson, "Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.

[2] M. Rußwurm and M. Körner, "Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification From Medium-Resolution Multi-Spectral Satellite Images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, Honolulu, HI, USA, July 2017, pp. 11–19.

[3] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series," *Remote Sens.*, vol. 11, no. 5, pp. 1–25, Mar. 2019.

[4] M. Rußwurm and M. Körner, "Self-attention for raw optical Satellite Time Series Classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 421–435, Nov. 2020.

[5] P. H. Swain, "Bayesian Classification in a Time-Varying Environment," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 12, pp. 879–883, Dec. 1978.

[6] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1350–1359, May 1999.

[7] L. Bruzzone, R. Cossu, and G. Vernazza, "Detection of land-cover transitions by combining multidate classifiers," *Pattern Recog. Lett.*, vol. 25, no. 13, pp. 1491–1500, Oct. 2004.

[8] N. Bouguila, W. Fan, and M. Amayri, *Hidden Markov Models and Applications*, Springer, Cham, Switzerland, 2022.

[9] D. Yu, L. Deng, and G. Dahl, "Roles of Pre-Training and Fine-Tuning in Context-Dependent DBN-HMMs for Real-World Speech Recognition," in *Proc. Neural Inform. Process. Syst. Worksh.*, Vancouver, Canada, Dec. 2010, pp. 1–8.

[10] F. Seide, G. Li, X. Chen, and D. Yu, "Feature engineering in Context-Dependent Deep Neural Networks for conversational speech transcription," in *Proc. IEEE Worksh. Autom. Speech Recog. Understanding*, Waikoloa, HI, USA, Mar. 2011, pp. 24–29.

[11] L. Li, Y. Zhao, D. Jiang, Y. Zhang, F. Wang, I. Gonzalez, E. Valentin, and H. Sahli, "Hybrid Deep Neural Network–Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition," in *Proc. IEEE Humaine Assoc. Conf. Affect. Comput. Intell. Interac.*, Geneva, Switzerland, Sep. 2013, pp. 312–317.

[12] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, "A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal," *Neurocomputing*, vol. 294, pp. 94–101, June 2018.

[13] S. J. Buchan, M. Duran, C. Rojas, J. Wuth, R. Mahu, K. M. Stafford, and N. Becerra Yoma, "An HMM-DNN-Based System for the Detection and Classification of Low-Frequency Acoustic Signals from Baleen Whales, Earthquakes, and Air Guns off Chile," *Remote Sens.*, vol. 15, no. 10, pp. 1–22, May 2023.

[14] R. Sedona, C. Paris, L. Tian, M. Riedel, and G. Cavallaro, "An Automatic Approach for the Production of a Time Series of Consistent Land-Cover Maps Based on Long-Short Term Memory," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Kuala Lumpur, Malaysia, July 2022, pp. 203–206.

[15] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Honolulu, HI, USA, July 2017, pp. 212–220.