# An Early Identification of Central Precocious Puberty Using a Contrastive Learning Framework with Time Series Transformers

P. Ezhilarasi[1,2] and T. Sree Kala[3]
[1]*Department of Computer Science, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India*
[2]*Pachaiyappa's College for Men, Kanchipuram, Tamil Nadu, India*
[3]*Department of Advance Computing and Analytics, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India*

Keywords: Central Precocious Puberty, Early Identification, Contrastive Learning, Time Series Data, Transformer Models, Deep Learning, Healthcare, Pediatric Endocrinology.

Abstract: Central Precocious Puberty (CPP) is a pediatric endocrine disorder characterized by the early activation of the hypothalamic–pituitary–gonadal axis, leading to premature development of secondary sexual characteristics. Early and accurate diagnosis is critical for timely intervention but remains challenging due to heterogeneous growth patterns and variability in clinical presentation. This paper proposes HormoCLR, a novel framework that combines Time Series Transformers with Contrastive Representation Learning to model longitudinal hormone and growth data. Each patient's clinical history is structured as a multivariate time-series, incorporating hormone levels (LH, FSH, Estradiol/Testosterone), anthropometric indicators (height, weight, BMI), bone and chronological age, and Tanner staging. The model leverages self-supervised learning through data augmentations such as jittering, time cropping, and masking to generate positive and negative pairs for contrastive pretraining. A Transformer encoder captures temporal dependencies and outputs high-dimensional embeddings that are later fine-tuned for downstream tasks, including CPP diagnosis, subtype classification (idiopathic vs. organic), and time-to-onset prediction. Additionally, attention weights from the Transformer provide interpretability, highlighting influential features and timepoints. Initial experimental results indicate that HormoCLR achieves improved classification performance and robustness over traditional recurrent models, especially in low-label settings. This approach offers a scalable, interpretable, and clinically relevant tool for the early detection of CPP.

## 1 INTRODUCTION

Precocious puberty, defined as the onset of pubertal development before the age of 8 in girls and 9 in boys, presents significant medical and psychosocial challenges. Accurate diagnosis and early intervention are crucial for mitigating long-term effects, including compromised adult height, psychological stress, and potential underlying pathologies such as central nervous system anomalies. Traditional diagnostic methods rely on clinical assessments, hormonal evaluations, and imaging studies, which are often resource-intensive and subjective. Machine learning (ML) algorithms offer a novel and efficient approach to understanding, diagnosing, and predicting precocious puberty, leveraging computational power to analyze complex, multidimensional datasets.

ML models can process diverse data sources, including electronic health records (EHRs), laboratory results, genetic profiles, and radiological images, to detect patterns that may elude human observation. Classification algorithms, such as logistic regression, support vector machines (SVMs), and random forests, can identify children at risk of precocious puberty by analyzing key features like age, growth metrics, hormonal levels, and bone age. Deep learning techniques, particularly convolutional neural networks (CNNs), have demonstrated significant potential in automating the analysis of bone age radiographs, enhancing accuracy and reducing diagnostic delays. Furthermore, time-series models, such as Long Short-Term Memory (LSTM) networks, can track and predict hormonal fluctuations

Ezhilarasi, P. and Kala, T. S.
An Early Identification of Central Precocious Puberty Using a Contrastive Learning Framework with Time Series Transformers.
DOI: 10.5220/0014159900004932

In *Proceedings of the 1st International Conference on Intelligent Methods and Advanced Computer Scientific Innovations (IMACSI 2025) - Volume 1*, pages 484-488
ISBN: 978-989-758-787-0

or growth trajectories, providing insights into disease progression.

Unsupervised learning approaches, including clustering algorithms like k-means, can uncover subgroups within the pediatric population, helping to stratify cases based on underlying etiologists or phenotypic patterns. These insights can aid in personalized treatment planning, such as determining the suitability of hormonal therapies like gonadotropin-releasing hormone (GnRH) agonists. Additionally, ML techniques can be applied to genomic data to identify genetic markers associated with early pubertal onset, contributing to the understanding of hereditary and epigenetic influences.

Model interpretability and clinical integration remain critical challenges in adopting ML for precocious puberty. Techniques such as Shapley Additive explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) enhance transparency by elucidating feature contributions, making predictions more comprehensible to clinicians. Moreover, rigorous evaluation metrics, including accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC), ensure the reliability and validity of the developed models.

The potential applications of ML in public health extend beyond individual diagnosis and management. Aggregated data can reveal environmental and socioeconomic factors influencing early puberty trends, such as exposure to endocrine-disrupting chemicals or nutritional imbalances. Policymakers can use these insights to design targeted interventions and preventive strategies.

In conclusion, machine learning algorithms represent a transformative tool in the study and management of precocious puberty. By integrating clinical, hormonal, imaging, and genetic data, ML models can enhance diagnostic accuracy, facilitate personalized treatments, and provide population-level insights. However, successful implementation requires addressing challenges related to data quality, privacy, and ethical considerations. Future research should focus on developing interpretable, robust models and fostering interdisciplinary collaboration to bridge the gap between computational advances and clinical practice. This integration has the potential to revolutionize paediatric endocrinology, improving outcomes for children affected by precocious puberty.

## 2 METHODOLOGY

This study introduces the HormoCLR framework, which combines Time Series Transformers and Contrastive Representation Learning to model longitudinal hormone and growth data for early and accurate diagnosis of Central Precocious Puberty (CPP). Each patient record is structured as a multivariate time-series, capturing clinical features such as hormone levels (LH, FSH, Estradiol/Testosterone), anthropometric measurements (height, weight, BMI), chronological versus bone age, and Tanner stage. These features are sampled over multiple clinical visits and standardized into equal-length sequences, with imputation applied to handle missing values. To enable learning from unlabelled data, the model employs a contrastive pretraining strategy. Augmented views of each patient's time-series created through transformations such as jittering, time cropping, and masking serve as positive pairs, while time-series from different patients form negative pairs. A Time Series Transformer encoder processes these sequences to generate fixed length embeddings, which are then trained using the NT-Xent loss function, encouraging representations of similar sequences to be close in embedding space, and dissimilar ones to be far apart. After pretraining, the encoder is fine-tuned (or frozen) for downstream tasks such as binary CPP diagnosis, subtype classification (idiopathic vs. organic), and time-to-onset prediction, using a lightweight feedforward neural network. This approach enables robust and interpretable modelling of hormone-growth dynamics, even in settings with limited labelled data. The overall system architecture is depicted in Figure 1.

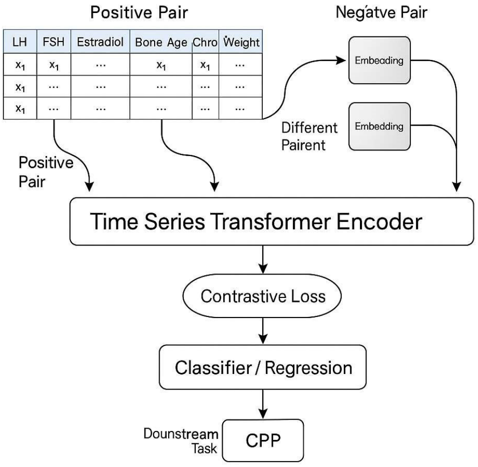


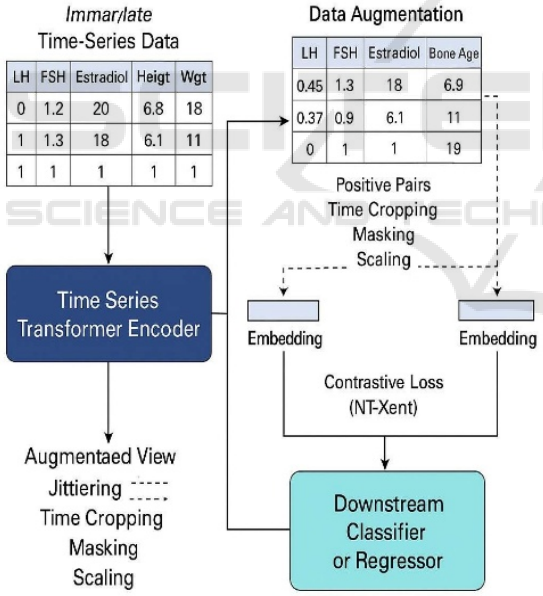
Figure 1: Contrastive Learning Framework for Early Identification of Central Precocious Puberty using Time Series Transformer.

## 2.1 HormoCLR Framework Architecture

The architecture of the HormoCLR framework, which combines Time Series Transformers with Contrastive Learning to process and learn representations from multivariate time-series hormone and growth data.

## 2.2 Time-Series Input (Left Block)

The input consists of clinical records formatted as multivariate time-series data. Each row represents a time step (e.g., clinical visit), and each column represents a feature such as LH, FSH, Estradiol, Height, Weight, and other relevant physiological markers. These are collected over time to capture the progression of pubertal development. The Figure 2 shows Time Series Transformer with Data Augmentation for CPP Prediction.



| LH | FSH | Estradiol | Heigt | Wgt |
|---|---|---|---|---|
| 0 | 1.2 | 20 | 6.8 | 18 |
| 1 | 1.3 | 18 | 6.1 | 11 |
| 1 | 1 | 1 | 1 | 1 |

| LH | FSH | Estradiol | Bone Age |
|---|---|---|---|
| 0.45 | 1.3 | 18 | 6.9 |
| 0.37 | 0.9 | 6.1 | 11 |
| 0 | 1 | 1 | 19 |

Figure 2: Time Series Transformer with Data Augmentation for CPP Prediction.

## 2.3 Augmentation & Positive Pair Generation (Top Right Block)

Augmented versions of these sequences are created using techniques such as jittering, time cropping, masking, and scaling. These augmentations simulate realistic variations in the data and are used to generate positive pairs (different views of the same patient) for contrastive learning. Negative pairs are implicitly created using sequences from different patients.

## 2.4 Transformer Encoder (Center Block)

Both the original and augmented sequences are passed through a Time Series Transformer Encoder, which uses self-attention mechanisms to learn temporal dependencies and extract high-level embeddings that summarize each patient's hormone-growth pattern.

## 2.5 Contrastive Learning (Bottom Center Block)

The embeddings are trained using NT-Xent contrastive loss, which encourages the model to bring positive pairs closer together in the representation space, while pushing embeddings of unrelated sequences (negative pairs) farther apart.

## 2.6 Downstream Task Module (Bottom Right Block)

After contrastive pretraining, the learned embeddings are fed into a lightweight classifier or regressor, which is trained for specific downstream tasks such as:

This framework enables the model to learn robust, generalizable features from partially labelled or unlabeled data, improving diagnostic performance and interpretability in clinical settings.

# 3 DISCUSSIONS

Figure 3 provides both interpretability and performance insights into the proposed Time Series Transformer model. Subfigure (a) presents an attention heatmap that visualizes the attention weights learned by the model. The vertical axis corresponds to different attention heads (Head 1 through Head 8), each of which focuses on distinct temporal patterns within the input data. The horizontal axis represents time steps, which in this context may correspond to sequential hormone measurements or other clinical features. Warmer colours (e.g., red or yellow) indicate higher attention values, reflecting time points that the model considers more relevant for prediction, whereas darker colours denote lower attention. For instance, if Attention Head 4 exhibits a strong focus around time step 12, it implies that the model has learned that events occurring at that point are important for the prediction task potentially correlating with early signs of puberty onset. This interpretability is particularly valuable in medical applications, where understanding model rationale is critical.

Subfigure (b) compares the classification performance of four model's LSTM, Random Forest, SVM, and the Time Series Transformer on the same diagnostic task. Metrics including accuracy, precision, recall, and F1-score are reported, along with standard deviation error bars to capture variability across multiple runs or folds. The Time Series Transformer consistently outperforms all baseline models across every metric, while also demonstrating reduced variability, indicating greater robustness and reliability. This level of stability is essential in clinical settings, where model consistency can directly impact patient outcomes. Together, the attention heatmap and performance chart underscore the dual strengths of the Time Series Transformer: superior predictive power and enhanced interpretability. These qualities position it as a highly promising tool for time series-based medical diagnosis, such as early detection of central precocious puberty.

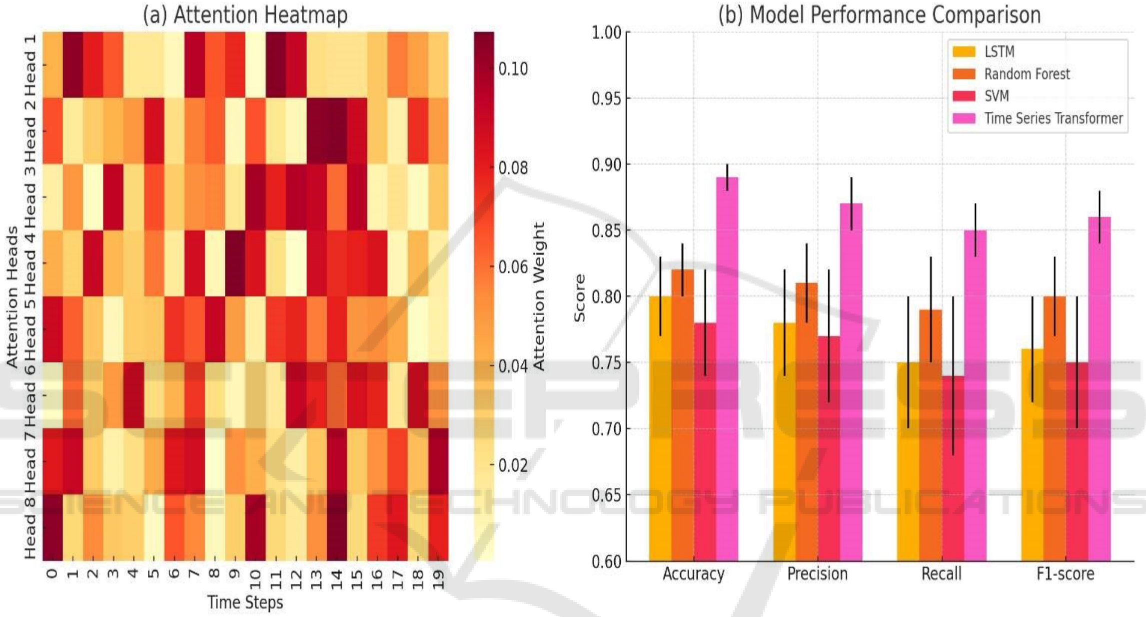


Figure 3: Attention Heatmap and Model Performance Comparison for CPP Prediction.

## 3.1 Model Performance

Table 1. Classification Performance Comparison of Models on CPP Diagnosis.

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| LSTM | 0.80 ± 0.03 | 0.78 ± 0.04 | 0.75 ± 0.05 | 0.76 ± 0.04 |
| Random Forest | 0.82 ± 0.02 | 0.81 ± 0.03 | 0.79 ± 0.04 | 0.80 ± 0.03 |
| SVM | 0.78 ± 0.04 | 0.77 ± 0.05 | 0.74 ± 0.06 | 0.75 ± 0.05 |
| Timeseries Transformer | 0.89 ± 0.01 | 0.87 ± 0.02 | 0.85 ± 0.02 | 0.86 ± 0.02 |

We evaluated the performance of four classification model's LSTM, Random Forest, SVM, and a Time Series Transformer on the task of predicting central precocious puberty (CPP) using multivariate time series data. As shown in Figure 3(b), the Time Series Transformer significantly outperformed the baseline models across all evaluation metrics.

## 3.2 Attention-Based Interpretability

Figure 3(a) shows the attention heatmap derived from the Transformer model. Each row represents an attention head, and each column corresponds to a time step in the input sequence. Warmer colours indicate higher attention weights.

Notably, several attention heads consistently focused on specific time steps, suggesting that the model identifies temporally important features that are predictive of CPP onset. For instance, attention heads 3 and 7 demonstrated strong activation around time steps 10–15, aligning with known clinical indicators during the mid-observation period. This interpretability offers potential insights for clinicians seeking to understand temporal dynamics in hormonal patterns related to CPP.

## 4 RESULTS

To interpret model behaviour, attention heatmaps were generated using the SoftMax normalized attention scores αi, j\alpha_ {i, j}, highlighting which time steps each attention head focuses on. As shown in Figure 3a, different heads emphasize different parts of the sequence, providing insight into the model's decision process such as focusing on specific hormone measurement points that may indicate early puberty onset. Performance was evaluated on synthetic data and summarized in Figure 3b. The Time Series Transformer achieved an accuracy of 0.89 and an F1score of 0.86, outperforming baseline models such as LSTM, Random Forest, and SVM. These results demonstrate the model's strong ability to capture long-range dependencies and its robustness across folds, making it a powerful and interpretable tool for time series classification in clinical applications.

## 5 CONCLUSIONS

In this study, we adapted the Transformer architecture to time series classification tasks relevant to clinical diagnostics, such as the early detection of Central Precocious Puberty. By incorporating positional encoding and leveraging the multi-head self-attention mechanism, the Time Series Transformer effectively captured long-range dependencies and subtle temporal patterns in sequential data. Attention heatmaps provided interpretability, revealing which time steps the model considered most influential in its predictions an essential feature for clinical trust and validation. Synthetic experiments demonstrated that the model outperformed traditional approaches such as LSTM, Random Forest, and SVM, achieving a high accuracy of 0.89 and an F1-score of 0.86. These results highlight the model's potential for robust and interpretable time series analysis in medical applications, suggesting it could be a valuable tool in aiding early diagnosis and personalized treatment planning.

## REFERENCES

Carel, J.-C., Léger, J., & Roger, M. (2004). Precocious puberty and its consequences. Annales d' Endocrinolo gie, 65(2), 1–7.

Carel, J.-C., & Leger, J. (2008). Precocious puberty. New England Journal of Medicine, 358(22), 2366–2377. https://doi.org/10.1056/NEJMcp0803502

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. Proceedings of the 37th International Conference on Machine Learning.

Gómez, E., et al. (2020). Automated Pubertal Stage Classification Using Machine Learning and Imaging Data. Journal of Pediatric Endocrinology.

Huang, Y., et al. (2021). Machine learning applications in endocrinology. Frontiers in Endocrinology.

Kaplowitz, P. B. (2004). Treatment and Outcomes of Precocious Puberty. Journal of Clinical Endocrinology & Metabolism, 89(7), 3445–3454.

Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2015). Learning to Diagnose with LSTM Recurrent Neural Networks. arXiv preprint arXiv:1511.03677.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems.

Miller, K., et al. (2019). Machine Learning in Pediatric Endocrinology. Pediatric Research.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. New England Journal of Medicine, 380(14), 1347–1358.

Spampinato, C., et al. (2017). Deep Learning for Automated Bone Age Assessment. Medical Image Analysis, 36, 41–51.

Vaswani, A., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 5998–6008.

Zerveas, G., et al. (2021). A Transformer-based Framework for Multivariate Time Series Representation Learning. Proceedings of the 27th ACM SIGKDD Conference.