A Multimodal Framework for Emotion Recognition in Children's Drawings Based on Machine Learning Techniques

Noor Khadour

Department of Computer Science, Bioengineering, Robotics and System Engineering (DIBRIS)
University of Genova

Supervisor
Gualtiero Volpe
Eleonora Ceccaldi

Co-Supervisors
Gabriele Romano
Nicola Corbellini

Laurea Magistrale in Computer Engineering - University of Genova

July 16, 2026

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Gualtiero Volpe, for his invaluable guidance, continuous support, and insightful feedback throughout the development of this thesis. His expertise and encouragement have been fundamental to the completion of this work.

I would also like to sincerely thank Professor Eleonora Ceccaldi for her guidance, constructive suggestions, and for providing the data and research support that significantly contributed to this study. Her knowledge and assistance have been greatly appreciated throughout the project.

I am grateful to my co-supervisors, Gabriele Romano and Nicola Corbellini, for their availability, technical support, and valuable discussions. Their assistance and recommendations have been instrumental in the implementation and refinement of the research presented in this thesis.

Finally, I am deeply thankful to my family and friends for their unconditional love, encouragement, and constant belief in me throughout my academic journey.

Dedication

To my Father Ali,

You believed in me when I doubted myself and gave me the courage to dream bigger than I ever thought possible. You encouraged me to step out of my comfort zone and pursue my studies in Italy, supporting me through every challenge, every decision, and every small detail along the way. I could not have achieved this milestone without your love, support, and sacrifice. I love you, Father.

To my Mother Fatat,

You were my first teacher in life, you taught me how to read and write and stood beside me throughout every step of my early education, guiding me with patience, love, and unwavering support. You saw your dreams reflected in mine, and hoped that one day I would earn a master's degree. Although pursuing this dream meant being far from home, I carried your love and encouragement with me every step of the way. I love you, Mother

To my little sister Nagham,

Watching you grow has been one of my greatest blessings, and your presence has always brightened even my most difficult days. No matter where life takes us, you will always hold a special place in my heart. I love you, and I am so proud to be your sister.

To my little brother Khder,

With every decision I make, I know I have my little brother behind me. You have always been, and continue to be, one of the strongest pillars in my life. I know how proud you are of me, and I hope you know how proud I am of you. Thank you for always standing by my side. I love you, Brother.

To my beloved husband, Majeed,

Before my dreams had a chance to take a flight, you believed in them. You saw a future I could only hope for, and with every word of encouragement, you gave me the strength to keep moving toward it. I still remember you telling me that one day these dreams would become my reality.

Years have passed, and here we are—in Italy, hand in hand, husband and wife—as you witness one of those dreams come true. Every step of this journey has been brighter because of your love.

As this dream finds its place among our most cherished memories, I look ahead with excitement to all the dreams still waiting for us. Together, we will continue to dream boldly, build fearlessly, and celebrate every new beginning.

You are my greatest blessing, my safest home, and my forever. I love you today, tomorrow, and for all the days to come.

This thesis is dedicated to all of you.

Abstract

Socio-emotional learning (SEL) plays a fundamental role in children's well-being, social development, and academic readiness. However, many existing digital educational tools rely primarily on screen-based interaction and offer limited support for embodied emotional engagement. Research in embodied cognition emphasizes that emotional experience is deeply grounded in bodily action, suggesting that effective emotional learning requires interaction modalities that integrate movement, perception, and expression. Although tangible user interfaces (TUIs) have begun to address this need by linking physical activity to digital representation, there remains a gap in the automatic interpretation of children's non-verbal emotional expressions, particularly in classroom contexts where real-time feedback and longitudinal reflection are desirable.

This thesis presents a machine learning–enhanced system designed to support embodied socio-emotional learning through a tangible drawing interface. Building upon the existing CambiaColore framework, the system captures children's drawing behavior using a physical tool equipped with retroreflective markers and tracked using an infrared camera. The resulting data include spatial trajectories, temporal dynamics, color selection, and drawing images, providing complementary sources for representing children's emotional expression. To automatically estimate affective states, three machine learning approaches are

investigated: a Random Forest Regressor trained on handcrafted trajectory features, a MobileNet convolutional neural network trained on drawing images, and a hybrid multimodal model that combines the predictions of both approaches to exploit the complementary information provided by the two modalities.

The models estimate children's affective states by mapping motion-derived and visual features onto the valence–arousal dimensions of the Russell Circumplex Model of Affect. Their predictive performance is compared through statistical analysis using the Wilcoxon Signed-Rank Test, enabling the assessment of statistically significant differences between the feature-based, image-based, and hybrid approaches. Experimental results demonstrate the effectiveness of the proposed framework and highlight the benefits of combining handcrafted trajectory features with deep visual representations for automatic affect prediction.

By combining handcrafted motion features with deep visual representations within a unified multimodal framework, this thesis demonstrates the feasibility of automatic emotion prediction from children's drawings in an embodied learning environment. Furthermore, the system enables the generation of structured summaries of emotional activity over time, supporting the creation of a collective emotional overview of classroom dynamics. This approach provides educators with a data-informed tool to facilitate emotional awareness, reflection, and discussion while preserving the embodied nature of children's interactions, thereby contributing to the development of intelligent educational technologies that support embodied socio-emotional learning.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The integration of socio-emotional learning (SEL) into formal education has received more attention in recent years due to mounting evidence that emotional competence is linked to long-term well-being, social functioning, and academic performance. Schools are now seen as crucial settings where children develop the emotional awareness and self-control needed for successful learning and social interaction, rather than only as places for cognitive development. The significance of incorporating emotional learning into regular classroom activities has been reinforced by large-scale studies and meta-analyses that show that SEL treatments can result in quantifiable gains in students' emotional abilities, behavior, and academic outcomes [1].

In parallel with this shift in education, advances in learning technology and human–computer interaction have enabled new tools that facilitate emotional expression. Specifically, studies in embodied cognition indicate that sensorimotor activity and body movement are intricately linked to emotional experience. Emotions are not just feelings that are experienced inwardly; they can also be expressed through gestures, physical activity, and interactions with the outside world. Even without explicit verbal labeling, it has been demonstrated that movement, drawing, and free-form motor expression can convey affective traits including arousal, valence, and intensity [2]. Children, for whom nonverbal and motor-based forms of communication frequently precede linguistic emotional articulation, benefit greatly from this viewpoint.

In this sense, interactive sketching systems are a promising medium for promoting emotional expression in educational contexts. Free drawing allows children to externalize their internal moods through motion, rhythm, pressure, and color selection, resulting in behavioral traces that can be studied as markers of affective processes. Previous research has shown that the dynamic features of hand-drawn lines, such as speed, acceleration, and fluidity, are associated with affective dimensions, particularly within the valence-arousal framework of emotion representation [3]. This framework offers a continuous and interpretable model of affect that avoids rigid category emotion labels and is ideal for expressing the subtle and transitory character of children's emotional states.

Despite these advances, most existing systems rely on qualitative interpretation or manual analysis of children's drawings, limiting their scalability and their ability to provide timely feedback in educational settings. Moreover, relatively few studies have explored multimodal machine learning approaches that jointly exploit drawing trajectories and visual representations for continuous affect estimation.

This thesis builds upon an existing tangible interactive platform designed to capture children's drawing trajectories during classroom activities. While the physical device and interaction design were created independently of this research, the system generates multimodal data describing the drawing activity, including spatial trajectories, temporal information, color usage, and images of the completed drawings. The primary scientific contribution of this thesis is the development of a multimodal machine learning framework for automatic affect prediction from children's drawings within the two-dimensional valence–arousal space.

To address this challenge, this thesis investigates three complementary machine learning approaches for continuous affect prediction from children's drawings. A Random Forest Regressor is trained on handcrafted trajectory features describing the dynamics of the drawing process, while a MobileNet convolutional neural network learns visual representations directly from the completed drawings. In addition, a hybrid multimodal model combines the predictions of both approaches to exploit the complementary information provided by motion-based and visual features.

Overall, this thesis contributes to the intersection of affective computing, embodied interaction, and educational technology by demonstrating the feasibility of multimodal machine learning for continuous emotion prediction from children's drawings. By producing continuous valence–arousal estimates rather than discrete emotion labels, the proposed framework aligns with contemporary SEL approaches that emphasize emotional awareness, reflection, and discussion over prescriptive classification. Ultimately, this work contributes to the development of emotionally aware technologies that are accessible, interpretable, and pedagogically grounded for classroom environments.

1.1 Motivations

The growing integration of socio-emotional learning (SEL) into formal education reflects the widespread recognition that emotional competencies are fundamental to children's academic performance, social functioning, and long-term well-being. Numerous studies have demonstrated that school-based SEL interventions can enhance children's emotional intelligence, attitudes, and behaviors in addition to improving their academic achievement [1]. Despite these advantages, SEL is nevertheless difficult to put into practice in classrooms, especially when it comes to developing emotional awareness in ways that are interesting, developmentally appropriate, and inclusive of a variety of learners.

The dependence of many current SEL activities on linguistic contemplation and abstract emotional labeling is a significant drawback. Language-based methods are useful, but they could not adequately take into account children's innate emotional expression styles, which are frequently centered on movement, body activity, and nonverbal cues. Embodied cognition research highlights how emotions are intimately related to sensory experiences and physical interactions with the environment, rather than being solely internal or symbolic constructions. According to this viewpoint, education including movement and physical interaction might offer more meaningful access to emotional experience, and learning about emotions is intrinsically embodied [4]. In early childhood education, where motor activity frequently precedes and promotes emotional and cognitive development, this is especially pertinent.

By allowing children to externalize emotions through physical action instead of just verbal reports, tangible user interfaces (TUIs) and embodied interaction systems present a possible solution to overcome these constraints. By integrating movement, visual expression, and social interaction in educational settings, tangible, drawing-based methods have been shown to improve socio-emotional learning [5]. Children can communicate emotions through dynamic aspects like speed, rhythm, and spatial organization through drawing activities in particular. This results in extensive behavioral traces that represent emotional states. Both individually and collectively, these traces can be digitally recorded and examined to encourage introspection, dialogue, and emotional awareness.

Even while these systems are good at helping children express their emotions, they frequently rely on self-report or teacher interpretation to interpret the drawings and movements of the children . This restricts scalability and adds subjectivity. Exploring automated techniques that can analyze expressive drawing activity in a principled and data-driven way is therefore highly motivated. By learning mappings between observable drawing dynamics and underlying emotive characteristics, developments in machine learning and affective computing offer a chance to close this gap. Children's gradual and changing emotional states are well-represented by dimensional models of emotion, such as the valence–arousal framework, which provide a continuous and interpretable depiction.

The primary objective of this thesis is to develop and evaluate a multimodal machine learning framework capable of inferring children's affective states from expressive drawing behavior captured through an existing tangible interaction platform. To this end, the thesis investigates three complementary approaches: a Random Forest Regressor based on handcrafted trajectory features, a MobileNet convolutional neural network trained on drawing images, and a hybrid model that combines both modalities. By mapping expressive movement and visual information onto the valence–arousal space, the proposed framework aims to support emotional awareness in a manner that is embodied, scalable, and compatible with classroom environments. Rather than replacing existing pedagogical practices, the proposed framework is intended to complement SEL activities by providing computational support for understanding children's emotional expression as it emerges through movement and drawing.

1.2 Context of the Study

This thesis is situated at the intersection of affective computing, embodied interaction, and socio-emotional learning (SEL) in educational settings. Schools are now widely acknowledged as the key setting for fostering children's emotional growth in addition to their intellectual education. This change is supported by a wealth of empirical data showing that structured SEL programs used in classroom environments can significantly enhance students' social conduct, emotional intelligence, and academic achievement [1]. As a result, there is an increasing need for educational resources that promote emotional awareness in ways that are developmentally appropriate and can be easily incorporated into regular classroom activities.

In this larger educational context, embodied and tactile learning systems have become potential tools for developing socio-emotional competencies. These systems, which are based on the ideas of embodied cognition, place a strong emphasis on the part that movement, physical interaction, and sensory experience play in emotional comprehension. In order to help children with SEL, recent research has suggested and assessed tactile drawing-based interfaces that let children express their feelings through movement, color, and artistic expression [5]. These systems are best suited for guided emotional learning rather than isolated solo use because they are usually used in classroom settings where teachers mediate exercises and encourage reflection.

The research described in this thesis expands on an existing tangible interface system that lets children draw freely with a real tool while having their movements monitored by a computer. Children's sketching behavior is described by the system's multimodal data, including spatial trajectories, temporal dynamics, color information, and images of the completed drawings. Crucially, the physical device and interaction paradigm design and implementation serve as the technological and experimental basis for the current investigation rather than being the main emphasis of this work. The availability of this multimodal dataset provides a unique opportunity to investigate the computational interpretation of expressive drawing behavior in terms of affective states.

From a research perspective, this thesis is further situated within studies that examine the relationship between children's expressive drawing behavior—including both motion characteristics and visual properties—and emotional expression. According to experimental data, kinematic characteristics of drawn strokes, including speed, acceleration, and fluency, can be mapped onto affective dimensions and exhibit a systematic variation with emotional intent [6]. Children's impromptu drawing activity in naturalistic educational environments has received relatively little attention, despite the fact that a large portion of previous research has concentrated on adult participants or controlled experimental settings. This disparity emphasizes the need for models that respect the expressiveness and diversity present in children's interactions while being able to generalize to actual classroom data.

In light of this, the present thesis focuses on the development and evaluation of a multimodal machine learning framework capable of inferring children's affective states from expressive drawing behavior captured through an embodied educational system. By grounding the proposed approaches in an authentic classroom-oriented platform, this work seeks to provide methodological contributions that are both scientifically rigorous and pedagogically relevant. At the same time, it recognizes the importance of developing machine learning solutions that adhere to ethical principles and are suitable for educational environments. The proposed framework is designed to support, rather than replace, educators by providing interpretable and data-informed insights into children's emotional expression while respecting the principles of responsible AI, including transparency, fairness, and the appropriate use of sensitive affective information. In this way, the thesis addresses not only the technical challenges of affect recognition but also the ethical and practical considerations associated with deploying affective computing technologies in real classroom settings.

1.3 Objectives and Contributions

In order to enhance the existing TRATTO tangible interface, the primary goal of this research is to develop and integrate an automatic affect prediction framework. While the existing system effectively captures children's expressive drawing behavior and transforms it into digital artifacts, it lacks a computational mechanism for interpreting this information in terms of affective states. This thesis addresses that limitation by proposing a multimodal machine learning framework capable of mapping children's expressive drawing behavior to continuous valence and arousal dimensions.

More specifically, this thesis investigates complementary machine learning approaches that analyze both handcrafted trajectory features and visual representations of children's drawings. Their integration enables the automatic estimation of affective states while preserving the embodied interaction paradigm of the original system.

The specific objectives:

- **Dataset Preparation and Annotation:** To curate, preprocess, and organize the multimodal dataset generated by the TRATTO platform, including drawing trajectories, color information, and drawing images, and to associate each sample with validated valence and arousal annotations.

- **Evaluation of Predictive Accuracy:** To evaluate and compare the predictive performance of the proposed models using regression metrics and statistical analysis, assessing their ability to estimate children's affective states within the valence–arousal space.

- **Model Deployment:** To deploy the selected affect prediction model within the existing Processing-based TRATTO environment, enabling automatic estimation of children's affective states following the completion of each drawing activity.

Contributions to the Research:

This thesis makes the following contributions to research and practice:

- **Multimodal Affect Prediction for Tangible Interfaces:**

  The study introduces an analytical layer to an existing tangible interface for socio-emotional learning, extending it from a purely expressive medium to a system capable of computational affect interpretation. This represents a step toward objective support for emotional labeling within educational activities.

- **Multimodal Machine Learning Framework:**

  This thesis proposes and evaluates a multimodal framework that combines handcrafted trajectory features with deep visual representations for continuous affect prediction from children's drawings.

- **Data-Supported Emotional Reflection:**

  The work enables the generation of structured summaries of emotional activity over time, forming the basis for a collective emotional overview that can support teacher-led reflection and discussion within the classroom.

- **Comparative Evaluation of Machine Learning Approaches:**

  The thesis provides an empirical comparison of feature-based, image-based, and multimodal approaches for predicting children's affective states, contributing insights into the strengths and limitations of each methodology for educational affective computing.

1.4 Overview of the Thesis

This thesis is organized into six chapters, each addressing a distinct phase of the research, from the theoretical foundations of socio-emotional learning to the development and evaluation of a multimodal machine learning framework for affect prediction. The chapters progressively introduce the background, methodology, implementation, experimental evaluation, and conclusions of the proposed work. A brief overview of each chapter is provided below.

Chapter 1 – Introduction

This chapter introduces the research context, motivation, and objectives of the thesis. It discusses the importance of socio-emotional learning (SEL), embodied interaction, and affective computing in educational settings, identifies the research gap, and presents the proposed multimodal machine learning framework developed to enhance the existing TRATTO tangible drawing platform through automatic affect prediction.

Chapter 2 – State of the Art

This chapter reviews the scientific literature relevant to the thesis. It first examines socio-emotional learning, embodied cognition, and tangible user interfaces in educational environments. It then discusses previous research on expressive drawing and movement analysis, followed by an overview of affective computing and machine learning techniques for emotion recognition, including both traditional machine learning methods and deep learning architectures.

Chapter 3 – System's Architecture

This chapter describes the TRATTO platform and its technical architecture. It presents the hardware and software components used to capture children's drawing behavior and explains how multimodal data—including drawing trajectories, temporal information, color selection, and drawing images—are acquired and prepared for subsequent analysis.

Chapter 4 – Methodology

This chapter describes the complete methodological framework adopted in this research. It details the dataset preparation and annotation process, the extraction of handcrafted trajectory features, the development of the Random Forest Regressor, the MobileNet convolutional neural network, and the hybrid multimodal model. It also presents the evaluation methodology, including the performance metrics and the statistical analysis used to compare the proposed approaches.

Chapter 5 – Results and Plotting

This chapter presents the experimental evaluation of the proposed models. It analyzes the performance of the Random Forest, MobileNet, and hybrid approaches for predicting children's affective states within the valence–arousal space and discusses the comparative results obtained through statistical analysis. Finally, it illustrates the practical implications of integrating automatic affect prediction into the CambiaColore platform.

Chapter 6 – Conclusions and Future Work

The final chapter summarizes the main contributions of the thesis and discusses their implications for affective computing, embodied interaction, and socio-emotional learning. It reflects on the limitations of the proposed framework and outlines potential directions for future research, including improvements to multimodal affect prediction models, broader classroom validation, and the continued development of ethically responsible AI technologies for educational environments.

Chapter 2

State of the Art

Summary

This chapter presented the theoretical background for this research. It discussed the importance of Social and Emotional Learning (SEL), embodied cognition, and Tangible User Interfaces (TUIs) in supporting children's emotional expression through interaction. It also reviewed how drawing behavior and movement characteristics can reflect emotional states, providing the basis for analyzing children's drawings. Finally, the chapter examined existing approaches to affective computing, including traditional machine learning, deep learning, and multimodal methods. Together, these topics provide the foundation for the methodology presented in the following chapters.

2.1 Social and Emotional Learning (SEL) in Modern Education

Social and Emotional Learning (SEL) has become an important part of modern education, recognizing that children's development extends beyond academic achievement. It focuses on helping children understand and manage their emotions, build positive relationships, and make responsible decisions, all of which contribute to both school success and long-term well-being [7][8].

2.1.1 The Evolution of SEL: From Theoretical Frameworks to Classroom Practice

SEL is rooted in developmental psychology, emotional intelligence, and social learning theory, which emphasize the importance of emotional competence for healthy social, behavioral, and psychological development [9][10]. These ideas gradually evolved into educational frameworks that support the integration of emotional learning into everyday classroom practice.

Among the most widely adopted frameworks is the one developed by the Collaborative for Academic, Social, and Emotional Learning (CASEL), which identifies five core competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making [11]. Numerous studies have shown that school-based SEL programs improve students' social-emotional skills, behavior, and academic performance, particularly when emotional learning is embedded within regular classroom activities rather than taught separately [12].

2.1.2 The RULER Approach and the Role of Emotional Labelling

The RULER approach is a well-established evidence-based framework for promoting emotional intelligence in educational settings. Developed at the Yale Center for Emotional Intelligence, it focuses on five core emotional skills: recognizing, understanding, labeling, expressing, and regulating emotions [13]. Among

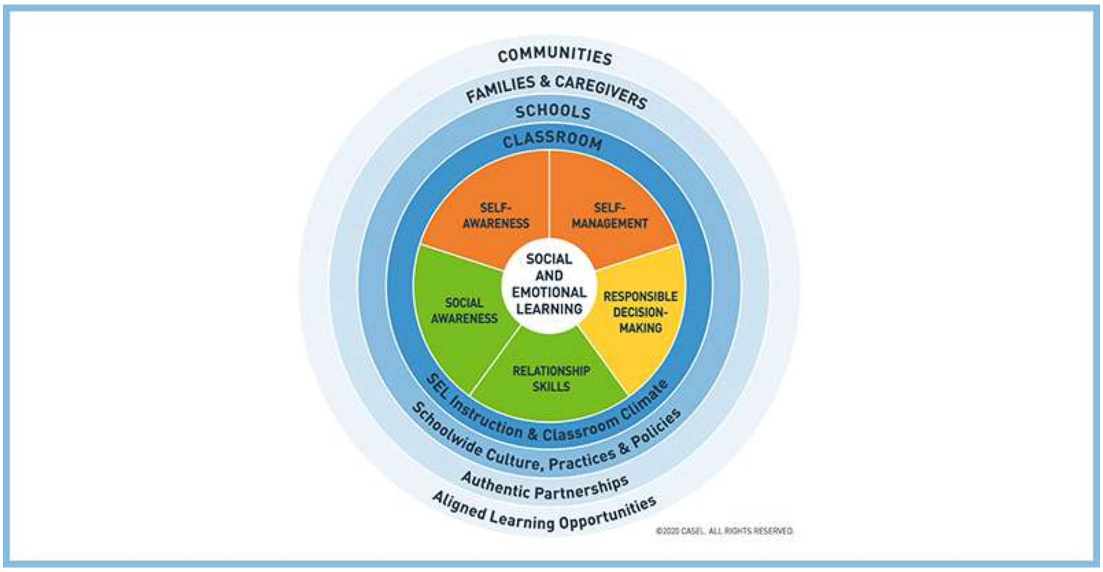


Figure 2.1: The CASEL Framework

these, emotional labeling plays a central role, as the ability to accurately identify and name emotions supports emotional regulation, communication, and positive social interactions [13][14].

RULER-based interventions have been shown to improve classroom climate, student engagement, and emotional well-being across preschool and primary school populations [15][16]. These findings highlight the importance of providing children with developmentally appropriate means of expressing and reflecting on their emotions. In this context, non-verbal forms of expression, such as drawing and movement, offer a complementary approach for supporting emotional awareness, particularly among younger children whose emotional vocabulary is still developing.

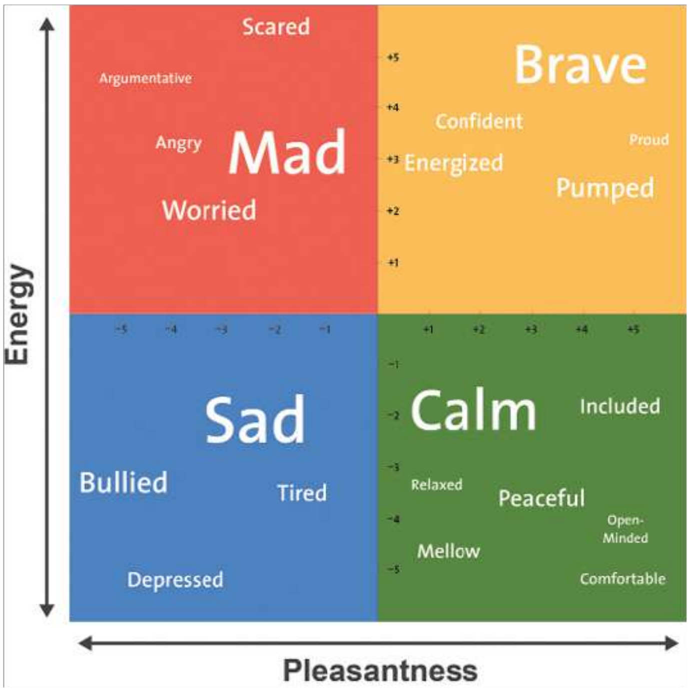


Figure 2.2: The RULER Mode Meter

2.1.3 Impact of Emotional Competence on Academic Readiness and Well-being

Research consistently shows that emotional competence is closely linked to children's academic readiness. Skills such as emotional regulation, impulse control, and stress management help children stay engaged in learning, overcome challenges, and build positive relationships with teachers and peers [17].

Emotional competence also supports long-term well-being. Studies have found that children who develop emotional awareness and regulation early are less likely to experience anxiety, behavioral difficulties, and social problems later in life [18]. These findings highlight the importance of integrating SEL into everyday school activities and motivate the development of innovative approaches, such as tangible interfaces and automated affect recognition, to support emotional learning in educational settings.

2.2 Theoretical Foundations of Embodied Emotion

Emotions are increasingly understood as embodied experiences shaped by the interaction between the body, the environment, and social context, rather than as purely cognitive processes. From this perspective, movement and drawing behavior can provide meaningful cues about affective states. This section presents the theoretical foundations that support embodied approaches to socio-emotional learning and affective computing.

2.2.1 The Embodied Mind: Cognition as an Enactive Process

Embodied cognition views thinking and emotion as processes shaped by continuous interaction between the body and the environment, rather than as purely internal mental activities. From this perspective, perception, action, and cognition are closely connected, and meaning is constructed through sensorimotor experience [19][20]. Emotions are therefore understood as being expressed through bodily actions as well as internal states.

The concept of the enactive mind further emphasizes that cognition develops through active engagement with the surrounding world [20]. This view has influenced educational research, where embodied learning approaches have been shown to enhance understanding, memory, and emotional engagement [21][22]. In this context, drawing and movement provide children with meaningful ways to express and reflect on their emotions, supporting both emotional learning and social interaction.

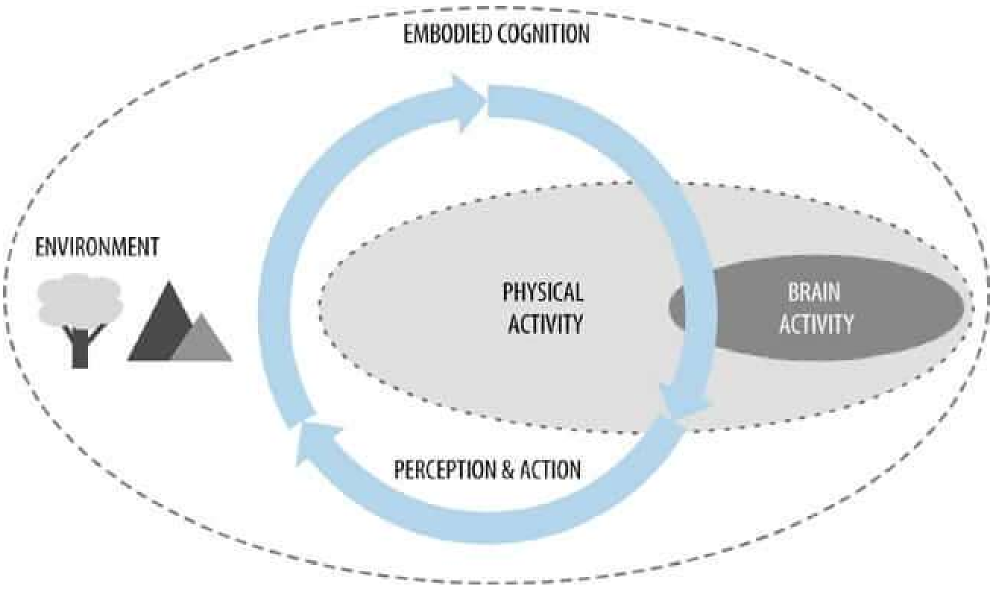


Figure 2.3: Embodied Cognition

2.2.2 The Relationship Between Motor Action and Affective Experience

Research in psychology and neuroscience has shown a close relationship between motor activity and emotional experience. Emotions are reflected in bodily activation and expressive movement, with characteristics such as speed, force, and fluidity varying according to emotional valence and arousal [23][24].

Studies on expressive movement have demonstrated that motion features extracted from dance, gestures, and drawing can communicate affective information [25][26]. In particular, features such as velocity, acceleration, rhythm, and smoothness have been linked to emotional intensity and polarity. Similar findings have been reported for hand-drawn lines, where the dynamics of drawing strokes influence emotional perception even in abstract drawings [6].

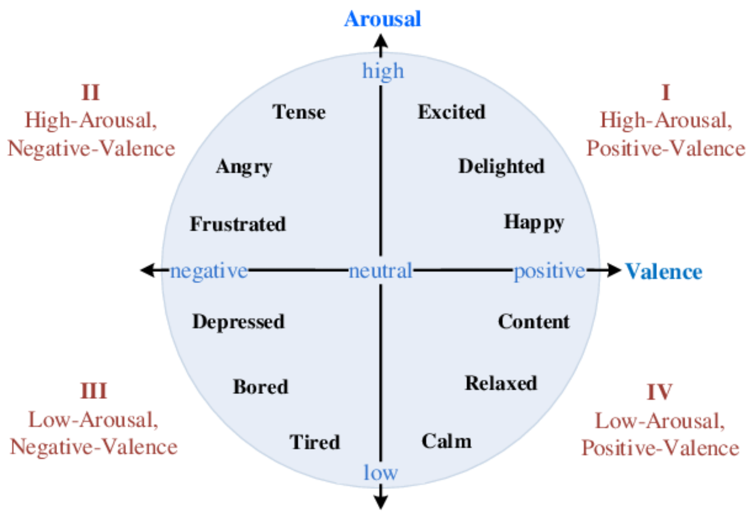


Figure 2.4: Affective Dimensions (Valence–Arousal)

These findings are especially relevant for children, who often express emotions through movement before they can describe them verbally. Consequently, drawing trajectories provide a valuable source of affective information and support the use of computational approaches to infer emotional states from observable behavior rather than relying solely on self-report.

2.3 Tangible User Interfaces (TUIs) for Learning

Tangible User Interfaces (TUIs) combine physical objects with digital information, allowing users to interact with technology through direct manipulation. In education, they have gained attention for supporting embodied learning, collaboration, and intuitive interaction, especially for children. By encouraging physical engagement, TUIs provide an effective alternative to traditional screen-based learning environments.

2.3.1 Defining Tangibility: Coupling Physical Tools with Digital Information

Tangible User Interfaces (TUIs) allow users to interact with digital information through physical objects instead of conventional input devices. In these systems, physical artifacts are part of the interface itself, making digital content more accessible through direct manipulation [27].

By linking physical interaction with digital information, TUIs support perception, action, and learning at the same time. This is particularly valuable in educational settings, where hands-on interaction can reduce cognitive load and help children better understand abstract concepts [28]. As a result, TUIs provide a natural bridge between physical experience and digital learning.

2.3.2 TUIs as Catalysts for Social Interaction and Group Conversation

Beyond individual interaction, TUIs are designed to support shared use, making them well suited for collaborative learning environments. By occupying a common physical space, they encourage face-to-face communication, joint attention, and interaction around shared artifacts, fostering richer social engagement than conventional screen-based interfaces [28].

Figure 2.5: TUIs as Social Catalysts

In educational settings, these social affordances promote discussion, collaboration, and emotional reflection among students and teachers [29]. Previous research has also shown that tangible systems can facilitate emotional communication by providing alternative, non-verbal means of expression, particularly for children [30][31]. Consequently, TUIs represent a valuable medium for supporting socio-emotional learning through shared interaction.

2.4 The Relationship Between Drawing Lines and Emotional Expression

Human beings have long used drawing as a way to express emotions that are difficult to communicate verbally. In children's drawings, emotional expression often appears through line characteristics such as curvature, rhythm, direction, and movement rather than realistic representation. Previous research has shown that these visual features can convey emotional meaning even in abstract drawings. This section reviews the main studies linking line qualities and drawing gestures to emotional expression.

2.4.1 Drawing Lines as Non-Verbal Emotional Expression

Research in experimental psychology has shown that lines can communicate emotions even without representing recognizable objects. Early studies found that people consistently associated sharp and angular lines with emotions such as anger, while smooth and curved lines were linked to calmer and more pleasant emotional states [35].

Later studies showed that observers, regardless of artistic experience, could reliably interpret the emotional meaning of abstract line drawings [36]. This suggests that emotional information is conveyed through perceptual features rather than learned symbols, making drawing an effective way for children to express emotions that they may not yet be able to describe verbally. Consequently, children's drawings provide valuable cues that can be observed and analyzed to better understand their emotional experiences.

2.4.2 Emotional Meaning in Line Qualities and Drawing Gestures

Later research has shown that the emotional meaning of a drawing depends not only on the final line but also on the gesture used to create it. Emotions influence movement characteristics such as speed, rhythm, and acceleration, leaving recognizable patterns in the resulting drawing. Early psychological studies suggested that lines appear expressive because they reflect the motor actions associated with emotional states [35].

More recent work has confirmed that observers associate implied movement, such as speed and force, with different levels of valence and arousal [37]. Similarly, motion-related features including velocity variation and acceleration have been shown to play an important role in conveying emotion through drawing [6]. These findings support the view that emotional meaning arises from the embodied drawing process, making movement dynamics a valuable source of information for computational affect recognition.

2.4.3 Interpreting Emotion from Abstract Drawing Traces

Research in psychology and affective computing suggests that abstract drawing traces contain meaningful emotional information, even without recognizable objects or symbols. Within dimensional models of emotion, such as the Circumplex Model of Affect, line characteristics have been consistently associated with valence and arousal. For example, smooth and continuous lines are generally linked to positive emotions, whereas angular and irregular lines tend to reflect negative emotions and higher arousal [36].

These findings provide a strong foundation for computational approaches that analyze children's drawings to estimate affective states. By learning patterns from drawing traces, machine learning models can support objective and scalable emotion recognition, complementing traditional observation in educational settings.

Chapter 3

System's Architecture

Summary

This chapter describes the technical architecture of the system developed in this research to support automated affect interpretation within the existing TRATTO tangible interface. While previous chapters established the theoretical foundations of socio-emotional learning, embodied interaction, and motion-based emotional expression, the present chapter focuses on the engineering design that enables the computational analysis of children's drawing behavior.

3.1 Overview of the System

Figure 3.1: TRATTO Tangible Drawing Interface

3.1.1 Tangible Drawing Interface (TRATTO)

TRATTO is a tangible drawing interface developed to capture children's drawing movements and convert them into digital trajectories in real time. The system enables users to interact with physical drawing tools while automatically recording their movements for later analysis.

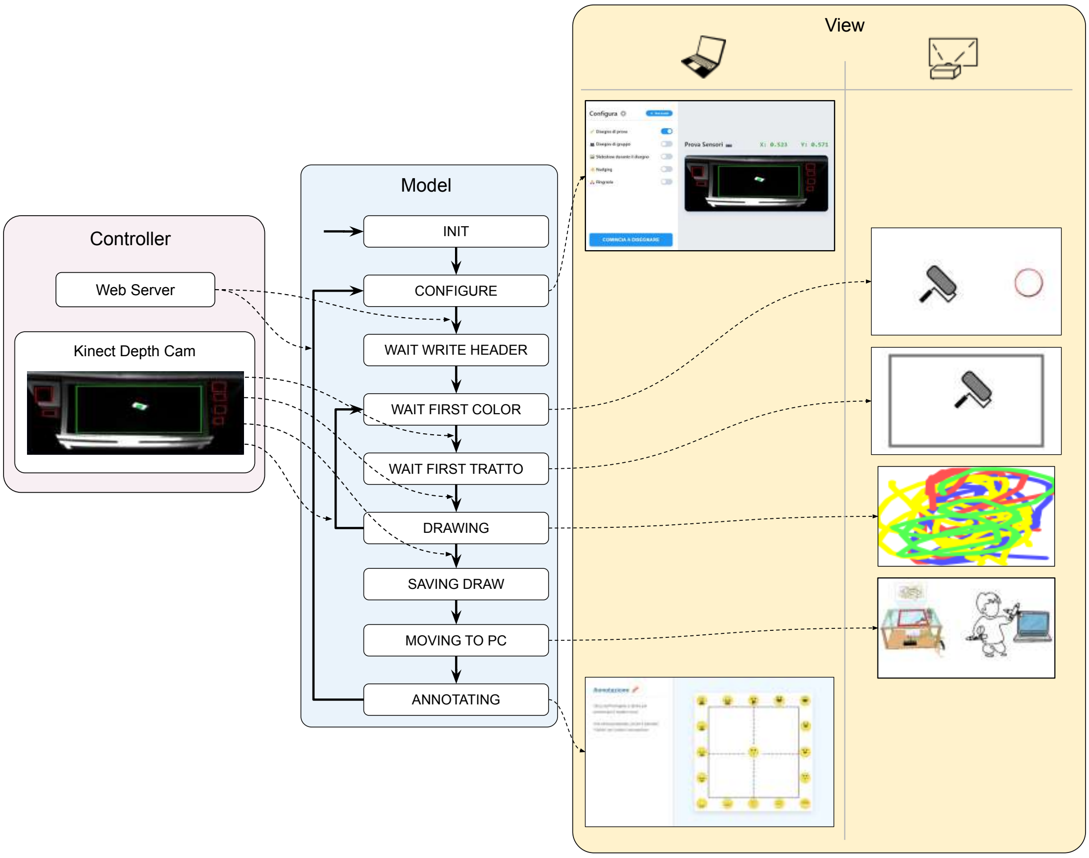


Figure 3.2: TRATTO System's Architecture

The platform consists of the following main components:

- A horizontal interactive table
- A projector mounted beneath the table
- A vertical projection surface supported by a tripod

- Physical drawing tools fitted with reflective motion markers

To interact with the system, the user holds a physical object, such as a stick, cup, roller, or glove, with a reflective marker attached to it. As the object moves across the interaction surface, an infrared camera tracks the marker's position, allowing the system to reconstruct the movement and display the corresponding drawing on the projected canvas in real time.

The interface provides several key functionalities, including:

- Real-time visual feedback
- Free-form drawing
- Color selection
- Automatic saving of drawing sessions

Unlike traditional drawing applications that rely on a mouse or touchscreen, TRATTO allows users to interact through natural hand movements using everyday physical objects. The drawing tools themselves contain no electronic components; they simply carry reflective markers that are detected by the tracking system.

During each drawing session, the interaction follows a simple sequence:

- Physical movement of the drawing tool
- Detection of the reflective marker
- Reconstruction of the drawing trajectory
- Real-time rendering of the drawing
- Storage of the recorded data

At the end of each session, the system produces a time-ordered record of the drawing trajectory, which serves as the basis for the feature extraction and emotion recognition methods presented later in this thesis.

The modular design of TRATTO also makes it possible to use different physical objects as drawing tools, provided they are equipped with a reflective marker. This flexibility allows the system to accommodate users with different motor abilities and makes it suitable for a variety of educational settings.

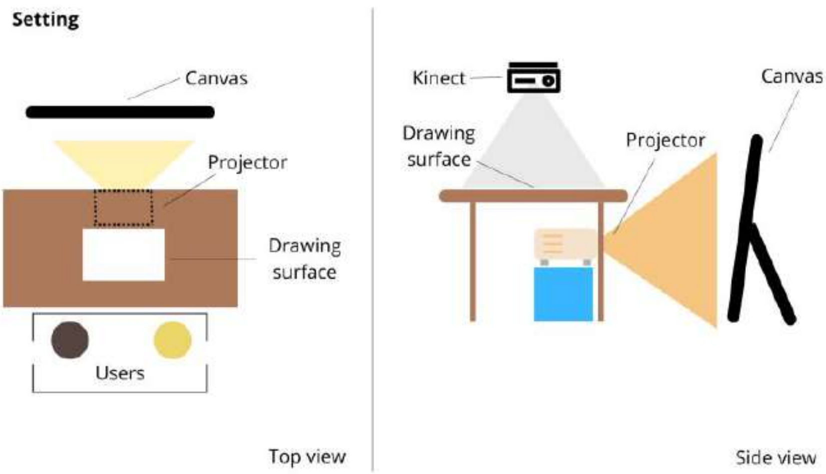


Figure 3.3: TRATTO Prototype

3.2 Data Acquisition and Collection Process

3.2.1 Experimental Drawing Protocol

The dataset consists of 130 drawings collected over two sessions. One session was carried out during a children's workshop at *La Città dei Bambini e dei Ragazzi* in Genoa, while the other one took place during the *Maker Faire* exhibition in Rome.

The drawing activity included two experimental conditions: **acted emotions** and **spontaneous emotions**. During each session, children selected a color and created a free-form drawing while the system tracked the movement of the drawing tool in real time. The recorded trajectory was automatically saved as a time-stamped file, and a rendered image of the completed drawing was stored as a PNG file.

Each drawing produced three synchronized outputs:

- A trajectory file containing the recorded movement data
- A rendered image of the drawing
- An annotation containing the corresponding emotional label

The same acquisition procedure was used in all sessions to ensure consistency across the dataset.

3.2.2 Acted and Spontaneous Emotional Conditions

The dataset includes two emotional drawing conditions: **acted emotions** and **spontaneous emotions**. The main difference between them is how the emotion is elicited during the drawing activity.

3.2.2.1 Acted Emotion Condition

In the acted condition, each child was asked to represent a specific emotion assigned by the experimenter. While the emotion was predefined, the children were free to decide how to express it through their drawings, including their

- Drawing movements
- Use of space

- Stroke patterns
- Choice of colors

Figure 3.4: example of an Acted Drawing

3.2.2.2 Spontaneous Emotion Condition

In the spontaneous condition, no emotion was assigned. Instead, children were invited to draw freely and express how they felt at that moment.

Unlike the acted condition, the emotion was not guided by the experimenter but reflected the child's natural emotional state during the activity.

Figure 3.5: example of a Non-Acted Drawing

3.2.3 Raw Data Generation

Each drawing session produces three synchronized files:

1. A trajectory log file (.txt)
2. A rendered drawing image (.png)
3. An annotation record (.tsv entry)

These files are linked using a unique filename, allowing each drawing to be associated with its trajectory and emotion annotation.

3.2.3.1 Trajectory Log File

The trajectory file contains the raw movement data recorded during a drawing session. Each row represents a single time sample and includes the timestamp, the marker position, its tracking status, and the selected drawing color.

The recorded fields are:

- `msec`, `sec`, `min`, `hour`: timestamp
- `x`, `y`: normalized coordinates
- `is_tracked`: tracking status (0 or 1)
- `color`: selected drawing color

The trajectory file serves as the input for the feature extraction process described in the next section.

25_10_17_trajectories_acted_g_10_id_1.txt - Notepad

File Edit Format View Help

msec	sec	min	hour	x	y	is_tracked	color
129	36	26	17	0.11551013	0.23592849	0	2
162	36	26	17	0.13088973	0.24649133	0	2
187	36	26	17	0.13088973	0.24649133	0	2
211	36	26	17	0.13088973	0.24649133	0	2
237	36	26	17	0.13088973	0.24649133	0	2
260	36	26	17	0.13088973	0.24649133	0	2
286	36	26	17	0.13088973	0.24649133	0	2
309	36	26	17	0.13088973	0.24649133	0	2
333	36	26	17	0.13088973	0.24649133	0	2
357	36	26	17	0.13088973	0.24649133	0	2
381	36	26	17	0.13088973	0.24649133	0	2
407	36	26	17	0.13088973	0.24649133	0	2
431	36	26	17	0.13088973	0.24649133	0	2
456	36	26	17	0.13088973	0.24649133	0	2
480	36	26	17	0.13088973	0.24649133	0	2
503	36	26	17	0.13088973	0.24649133	0	2
525	36	26	17	0.13088973	0.24649133	0	2

Figure 3.6: Trajectories example

3.2.3.2 Rendered Drawing Image

For each drawing session, the system also saves a rendered image in PNG format. The image is a visual representation of the drawing produced from the recorded trajectory.

Each image is linked to its corresponding trajectory file through a shared filename, which includes the acquisition date, group ID, participant ID, and drawing condition (acted or spontaneous).

Figure 3.7: Drawing example

3.2.3.3 Annotation Record

Each drawing is also recorded in an annotation file (`.tsv`), which stores the metadata for that session. This includes the acquisition date, group and participant IDs, the paths to the trajectory and drawing files, and the normalized valence (`x_normalized`) and arousal (`y_normalized`) coordinates provided by the child.

The valence and arousal values represent the child's self-reported emotional state after completing the drawing.

3.2.4 Data Curation and Cleansing

Before feature extraction, the dataset was manually reviewed to remove drawings that were not suitable for analysis. While most drawings consisted of free-form movement, some contained recognizable objects or scenes, making them less representative of movement-based emotional expression.

3.2.4.1 Exclusion Criteria

A drawing was excluded if it:

- relied mainly on recognizable symbols or objects,
- represented a staged or acted scene
- contained little usable movement data due to weak or noisy trajectories.

These drawings were removed to reduce the influence of visual content on the analysis.

Figure 3.8: Not Accepted Drawing Example

3.2.4.2 Inclusion Criteria

A drawing was retained if it:

- reflected natural drawing movements
- expressed emotion through the drawing process rather than recognizable symbols
- contained sufficient movement variability for feature extraction.

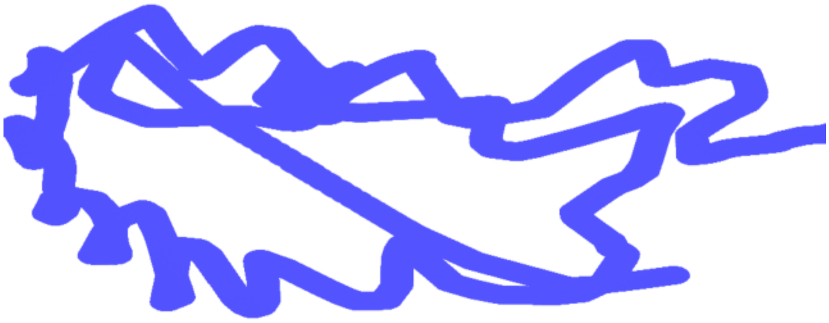


Figure 3.9: Accepted Drawing Example

3.3 Feature Extraction

This part describes the features extracted from the recorded drawing trajectories. The aim is to capture different aspects of the drawing process, including its temporal, kinematic, geometric, and spectral characteristics. In total, 25 features were extracted, with each feature representing a different property of the user's drawing behavior.

3.3.1 Visual Features

3.3.1.1 First Color

The *First Color* feature records the first color selected by the child from the four available options: blue, yellow, red, and green. As this choice is made at the beginning of the drawing activity, it may provide useful information about the child's initial approach to the task and expressive behavior.

Although the meaning of individual colors varies between individuals, the first color choice can reflect personal preferences or drawing habits. For this reason, it is included as a visual feature that captures an early aspect of the drawing process.

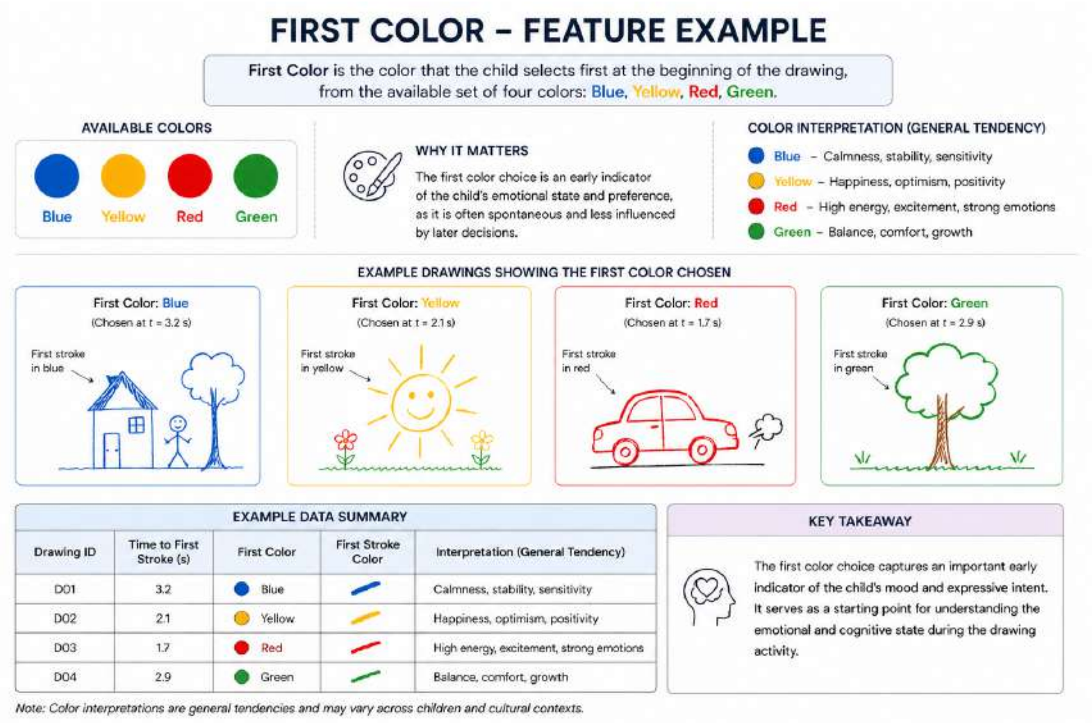


Figure 3.10: First Color Feature

3.3.2 Temporal Features

3.3.2.1 Time on Paper

Time on Paper measures the total time the drawing tool remains in contact with the drawing surface during a session. This feature reflects how long the child actively engages in the drawing task. Longer drawing times may indicate greater focus, carefulness, or emotional involvement, while shorter times can be associated with quicker or more spontaneous drawing behavior. As a result, this feature provides useful information about the child's interaction with the drawing activity.

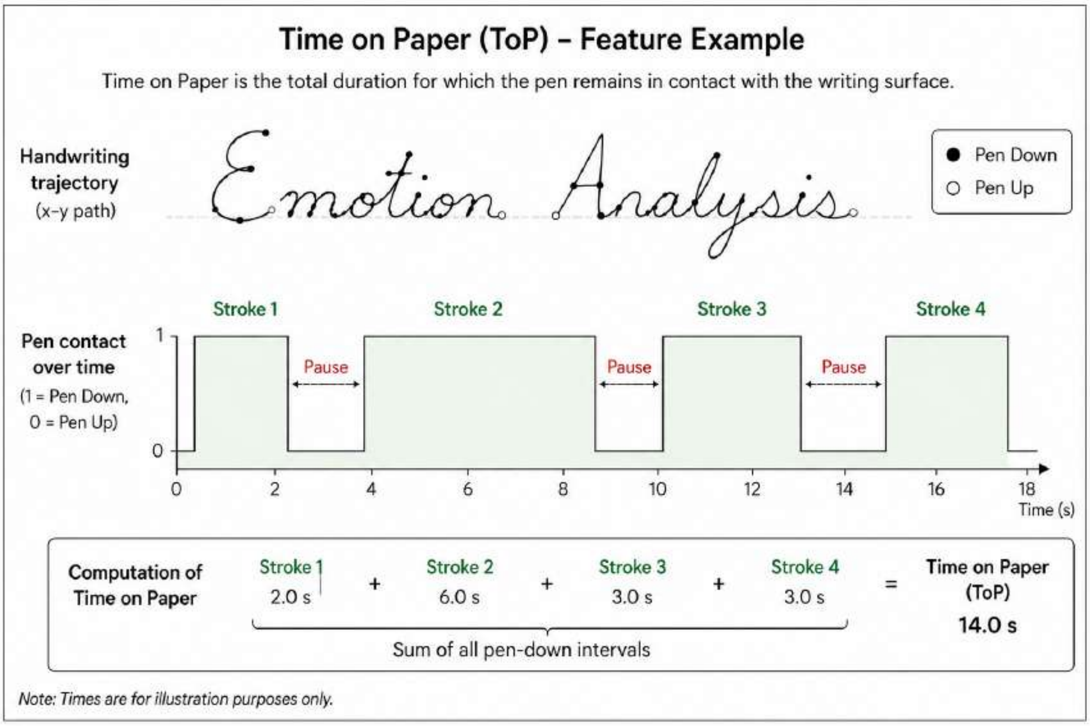


Figure 3.11: Time On Paper Feature

3.3.2.2 Stroke Duration

Stroke Duration measures the time taken to complete a single continuous stroke, from when the drawing tool touches the surface until it is lifted. This feature reflects how the child performs individual drawing movements. Longer strokes may indicate more careful or controlled movements, while shorter strokes are often associated with quicker and more spontaneous drawing behavior. As such, stroke duration provides information about the child's drawing style and motor control.

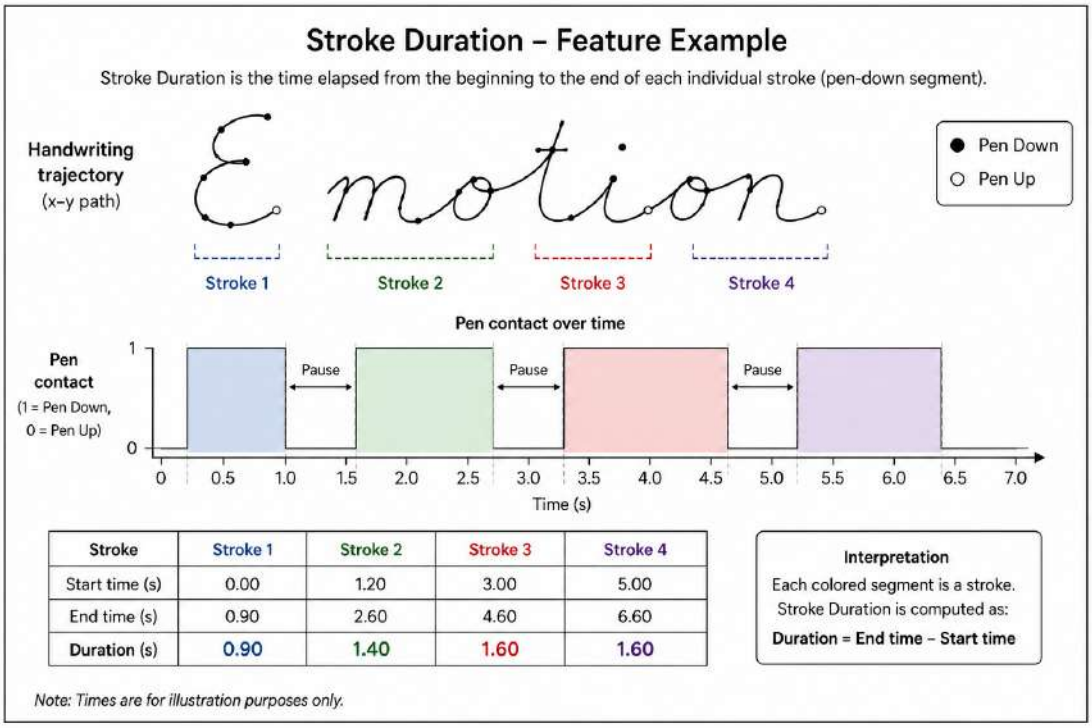


Figure 3.12: Stroke Duration Feature

3.3.3 Kinematic Features

3.3.3.1 Speed

Speed measures how quickly the drawing tool moves across the surface during a stroke. It reflects the pace of the child's drawing movements and can provide information about their drawing behavior. Higher speeds are generally associated with faster and more energetic movements, while lower speeds may indicate slower, more controlled drawing. As a result, speed is a useful feature for characterizing the dynamics of the drawing process.

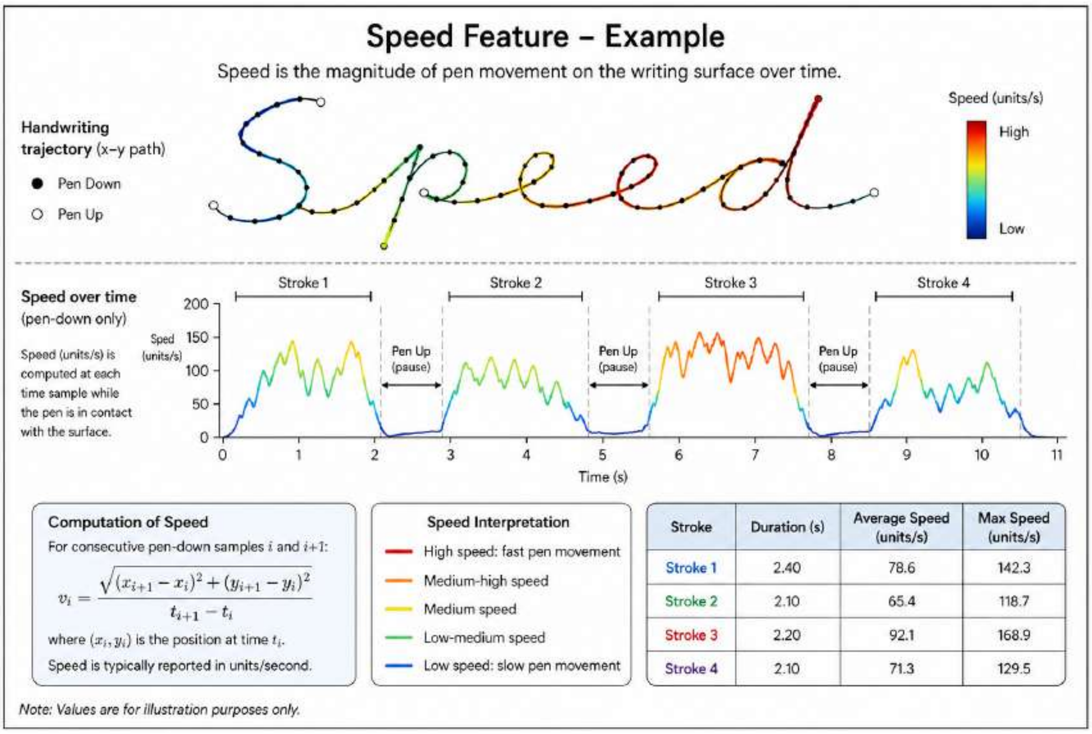


Figure 3.13: Speed Feature

3.3.3.2 Acceleration

Acceleration measures how quickly the drawing speed changes during a stroke. It reflects variations in the child's drawing movements over time. Higher acceleration indicates more abrupt changes in speed, suggesting faster or less controlled movements, while lower acceleration is associated with smoother and more consistent motion. This feature helps characterize the dynamics and stability of the drawing process.

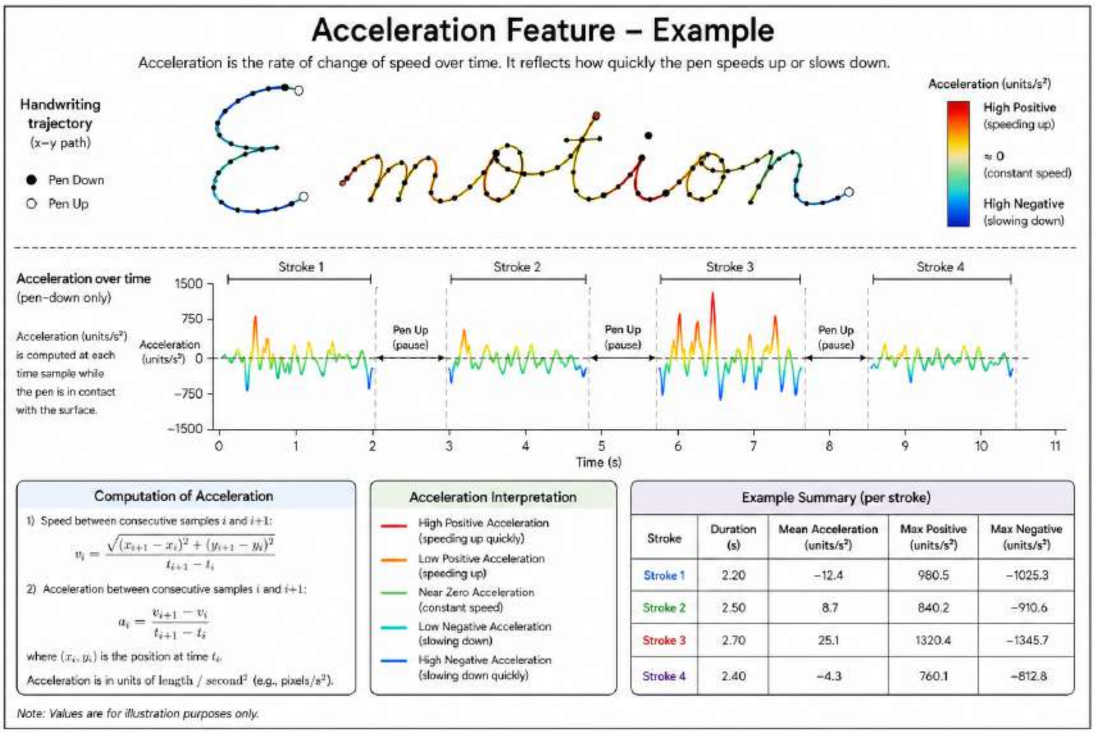


Figure 3.14: Acceleration Feature

3.3.3.3 Velocity Variation

Velocity Variation measures how much the drawing speed changes throughout a stroke. It reflects the consistency of the child's drawing movements. Higher values indicate greater fluctuations in speed, while lower values correspond to smoother and more uniform motion. This feature helps describe the stability and dynamics of the drawing process.

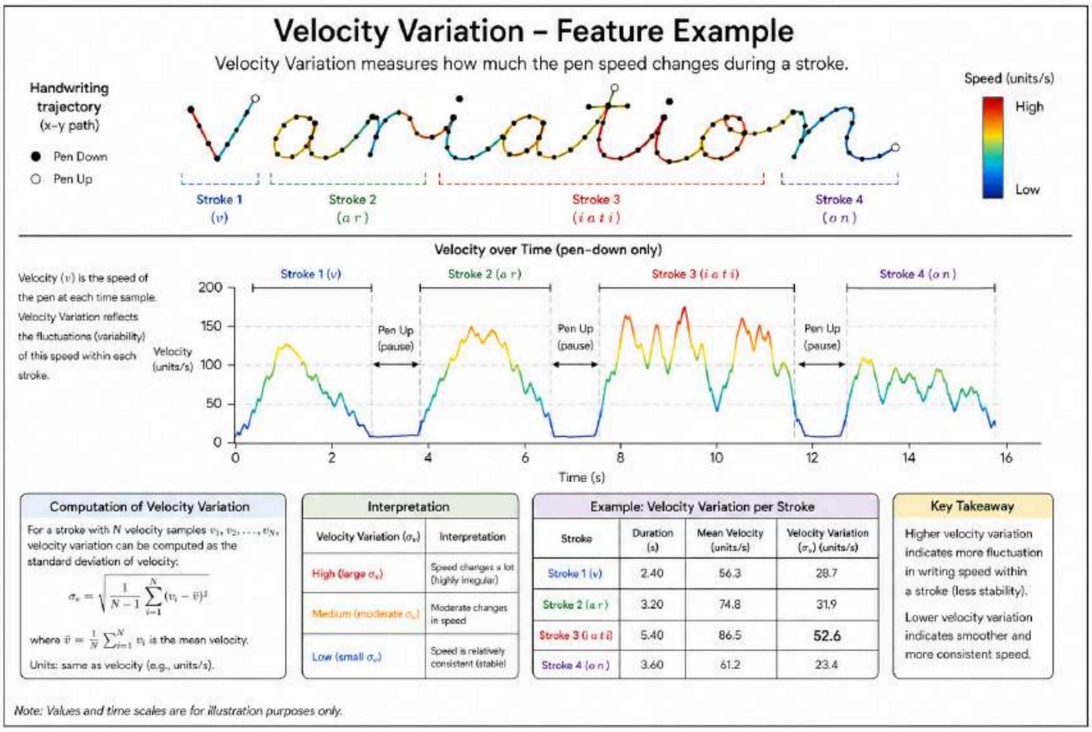


Figure 3.15: Velocity Variation Feature

3.3.3.4 Y Velocity Energy Focus

Y Velocity Energy Focus measures how much of the drawing movement is concentrated along the vertical (Y) axis. Higher values indicate stronger up-and-down movements, while lower values suggest a more balanced or horizontal drawing pattern. This feature captures the dominant direction of movement and helps characterize the child's drawing behavior.

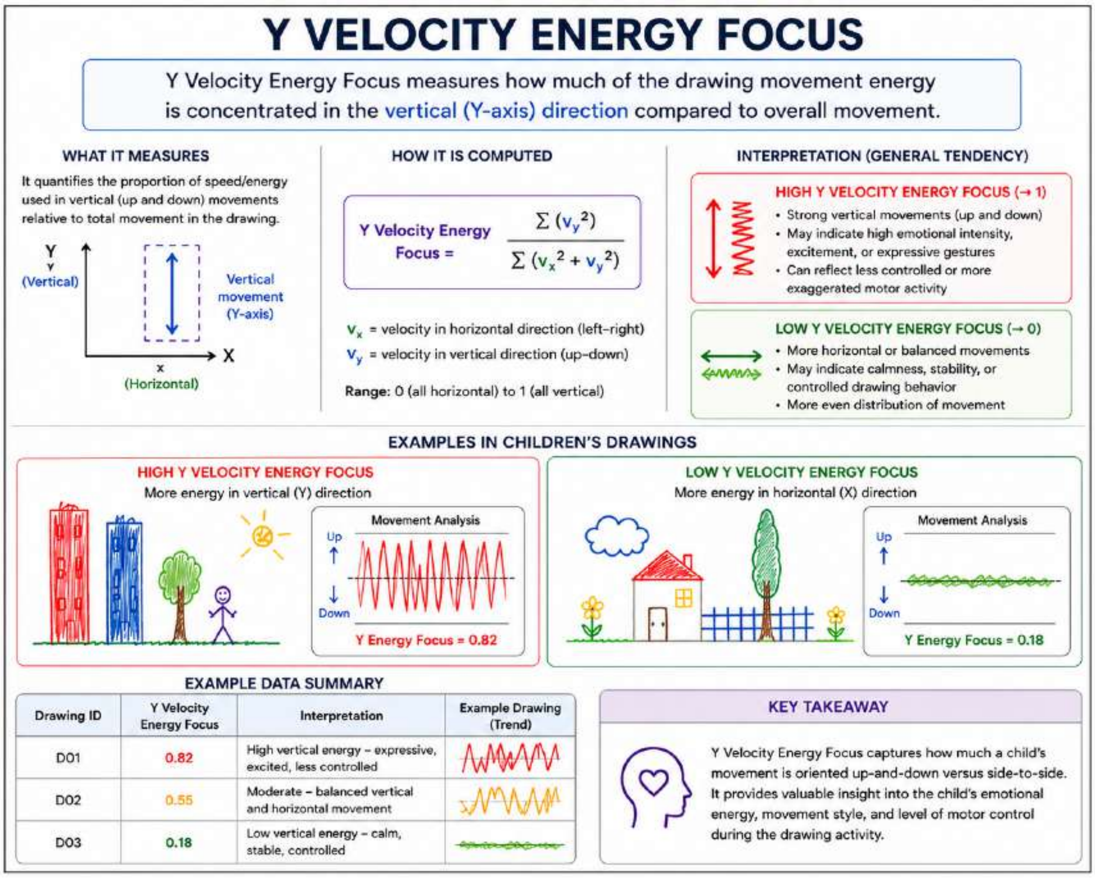


Figure 3.16: Y Velocity Energy Focus Feature

3.3.4 Statistical / Distribution Features

3.3.4.1 Y Average

Y Average represents the mean vertical (Y-axis) position of all recorded points in a drawing. It indicates where the drawing is generally located along the vertical axis. Higher values correspond to drawings concentrated toward the upper part of the drawing area, while lower values indicate drawings positioned closer to the bottom. This feature describes the child's overall use of vertical space during the drawing process.

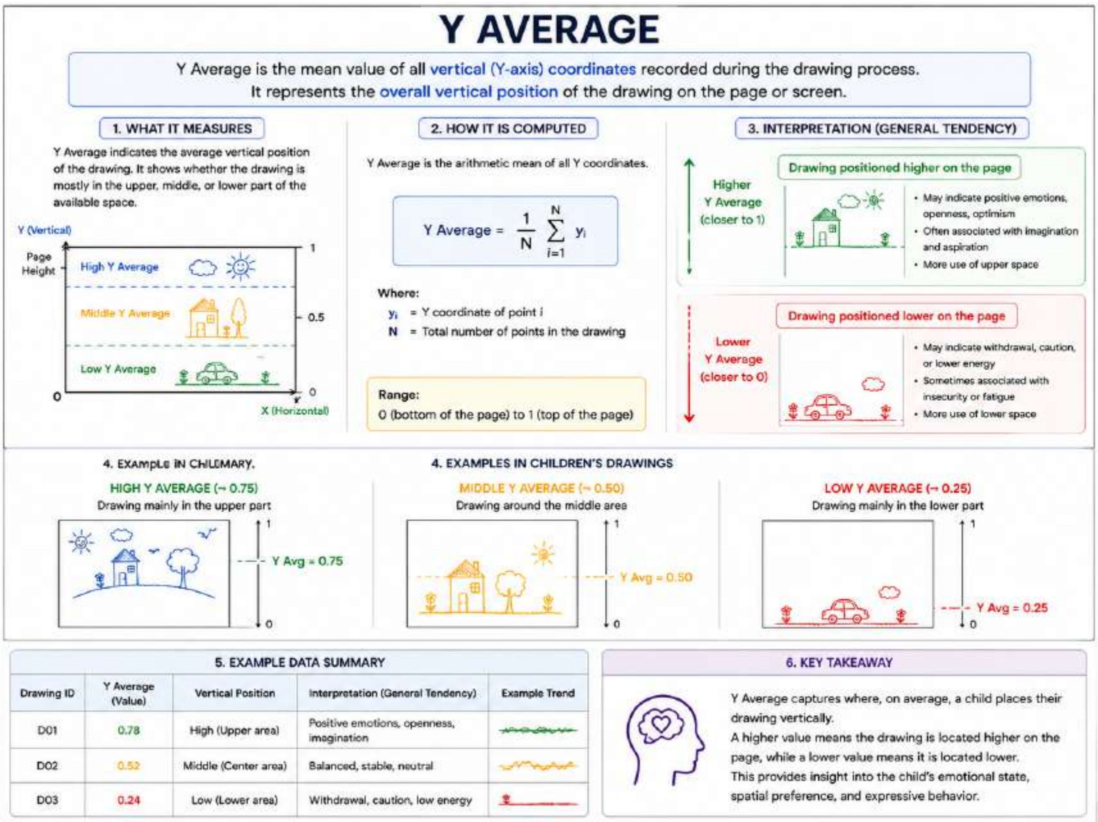


Figure 3.17: Y Average Feature

3.3.4.2 Distribution Symmetry

Distribution Symmetry measures how evenly the drawing is distributed around a central axis. Higher symmetry indicates a more balanced and organized drawing, while lower symmetry reflects a more uneven or irregular distribution. This feature describes the overall spatial organization of the drawing and the child's use of the drawing space.

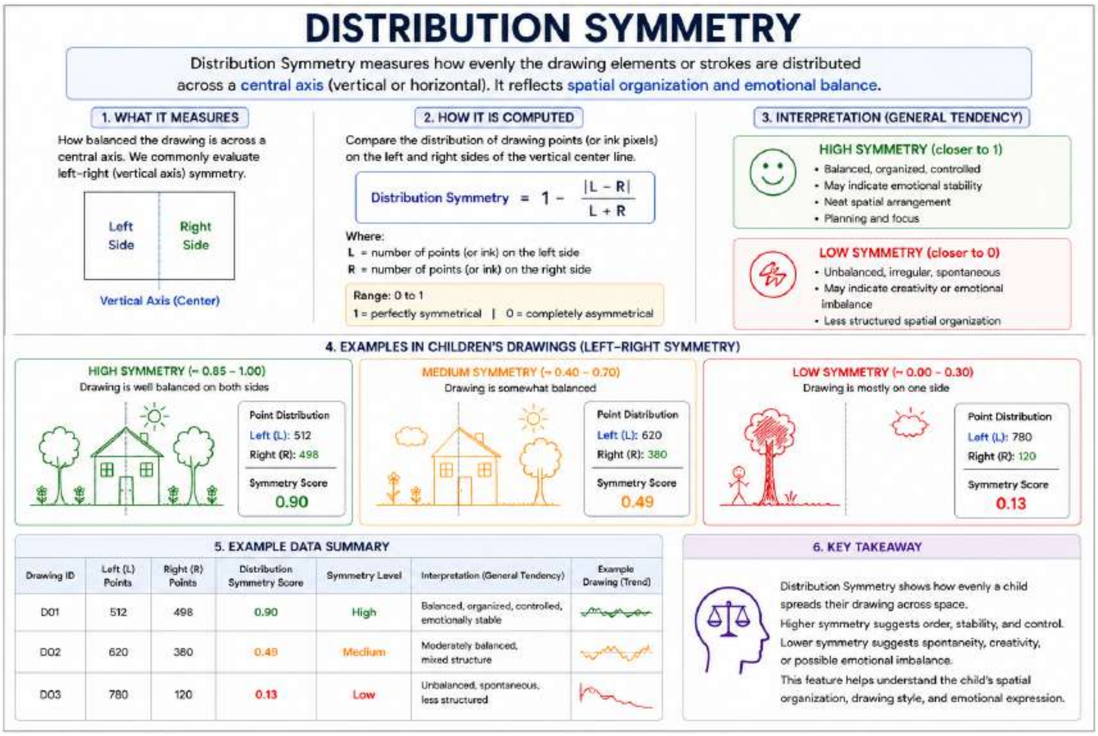


Figure 3.18: Distribution Symmetry Feature

3.3.5 Structural / Behavioral Features

3.3.5.1 Number of Strokes

Number of Strokes represents the total number of continuous strokes used to complete a drawing. A stroke begins when the drawing tool touches the surface and ends when it is lifted. Higher values indicate that the drawing was created using many separate strokes, while lower values correspond to fewer, longer strokes. This feature captures the child's drawing strategy and the overall structure of the drawing process.

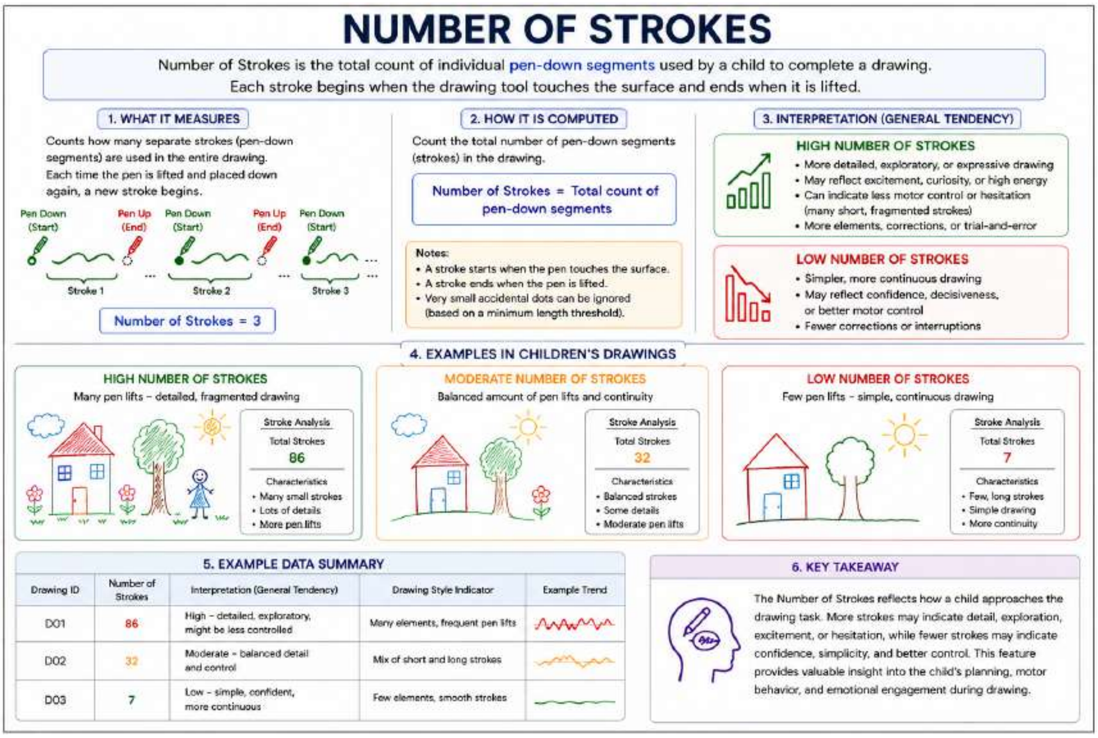


Figure 3.19: Number of Strokes Feature

3.3.5.2 Number of Inflexions

Number of Inflexions measures how many times the curvature of a drawing changes direction along a stroke. Higher values indicate more frequent changes in the drawing path, resulting in more complex or irregular shapes, while lower values correspond to smoother and simpler curves. This feature describes the complexity and variability of the child's drawing movements.

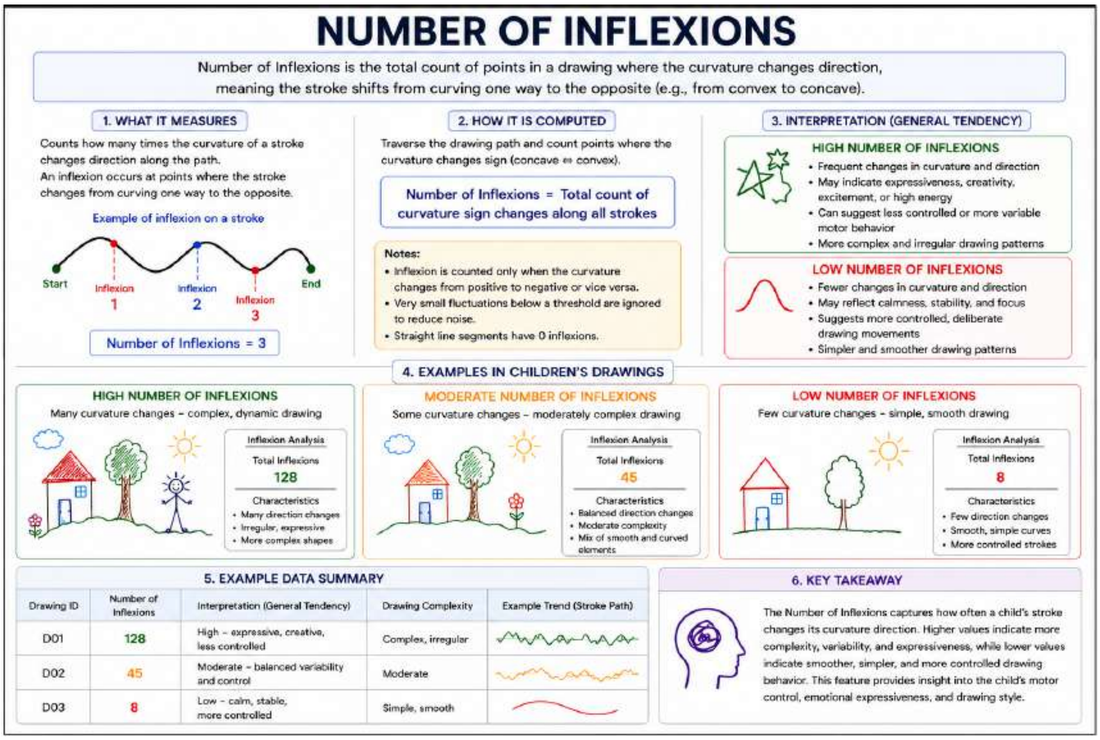


Figure 3.20: Number of Inflexions Feature

3.3.5.3 Closure

Closure measures the extent to which a drawing contains closed shapes by evaluating whether the start and end points of strokes connect. Higher values indicate more complete and enclosed shapes, while lower values correspond to more open or incomplete drawings. This feature describes the child's ability to produce connected and well-defined drawing structures.

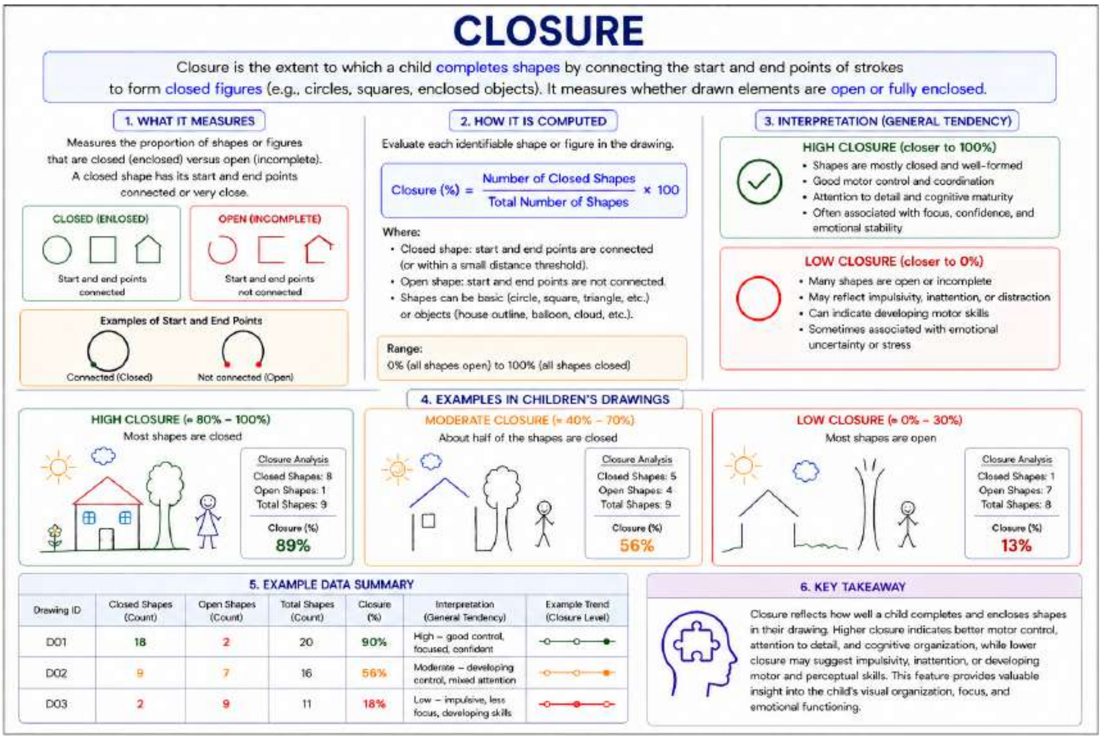
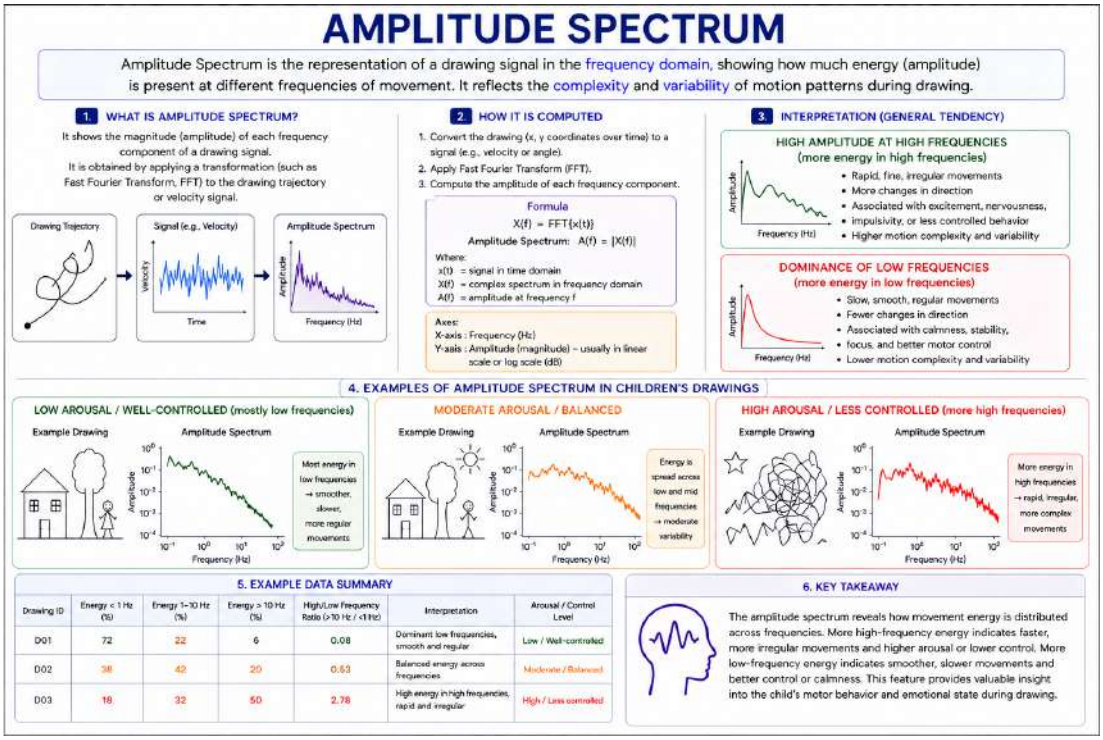


Figure 3.21: Closure Feature

3.3.6 Frequency / Spectral Features

3.3.6.1 Amplitude Spectrum

Amplitude Spectrum represents the distribution of movement energy across different frequencies in the drawing signal. It is computed by applying the Fourier Transform to the recorded trajectory. Higher amplitudes at higher frequencies indicate more rapid and irregular movements, while lower-frequency components correspond to smoother and more regular motion. This feature captures the frequency characteristics of the child's drawing behavior.



Drawing ID	Energy < 1 Hz (%)	Energy 1–10 Hz (%)	Energy > 10 Hz (%)	High/Low Frequency Ratio (>10 Hz / <1 Hz)	Interpretation	Arousal / Control Level
D01	72	22	6	0.08	Dominant low frequencies, smooth and regular	Low / Well-controlled
D02	38	42	20	0.53	Balanced energy across frequencies	Moderate / Balanced
D03	18	32	50	2.78	High energy in high frequencies, rapid and irregular	High / Less controlled

Figure 3.22: Amplitude Spectrum Feature

3.3.6.2 Frequency Centroid

Frequency Centroid represents the average frequency of the drawing signal, weighted by the amplitude spectrum. It indicates whether the movement is dominated by lower or higher frequency components. Higher values correspond to faster and more irregular movements, while lower values indicate smoother and more gradual motion. This feature summarizes the overall frequency characteristics of the child's drawing behavior.

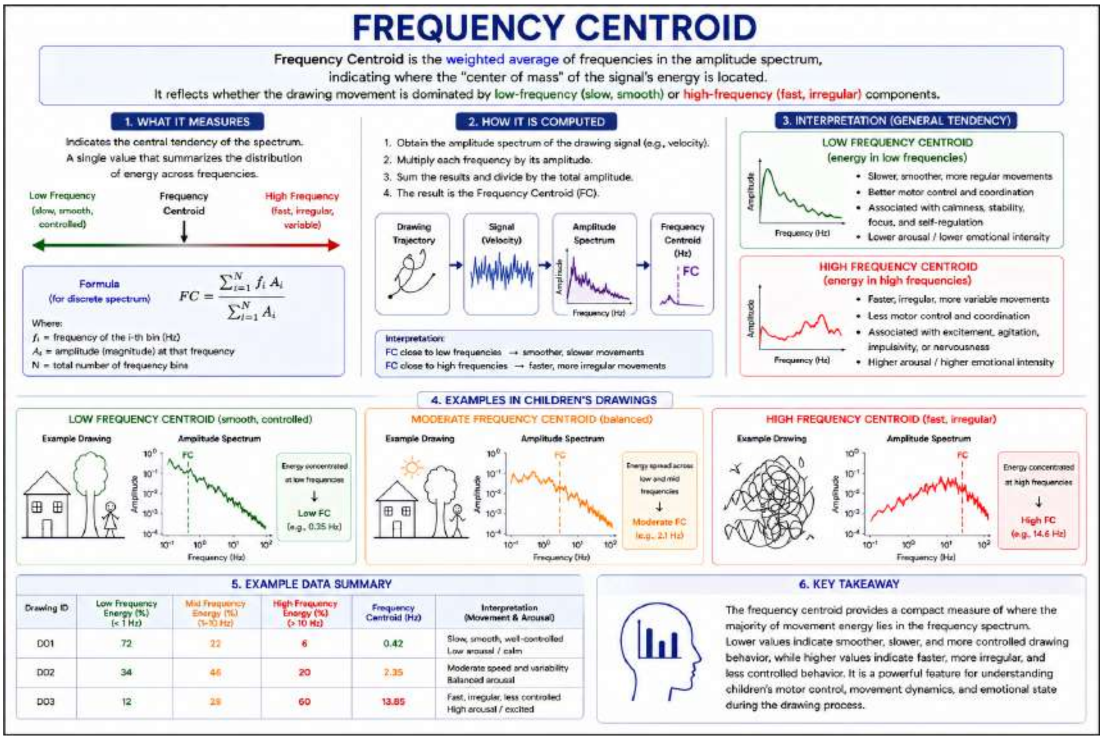


Figure 3.23: Frequency Centroid Feature

3.3.6.3 Spectral Features

Spectral Features are extracted from the frequency-domain representation of the drawing signal after applying the Fourier Transform. Unlike time-based features, they describe how movement energy is distributed across different frequencies. Together, these features capture the rhythm, variability, and overall frequency characteristics of the drawing movements, providing complementary information about the child's drawing behavior.

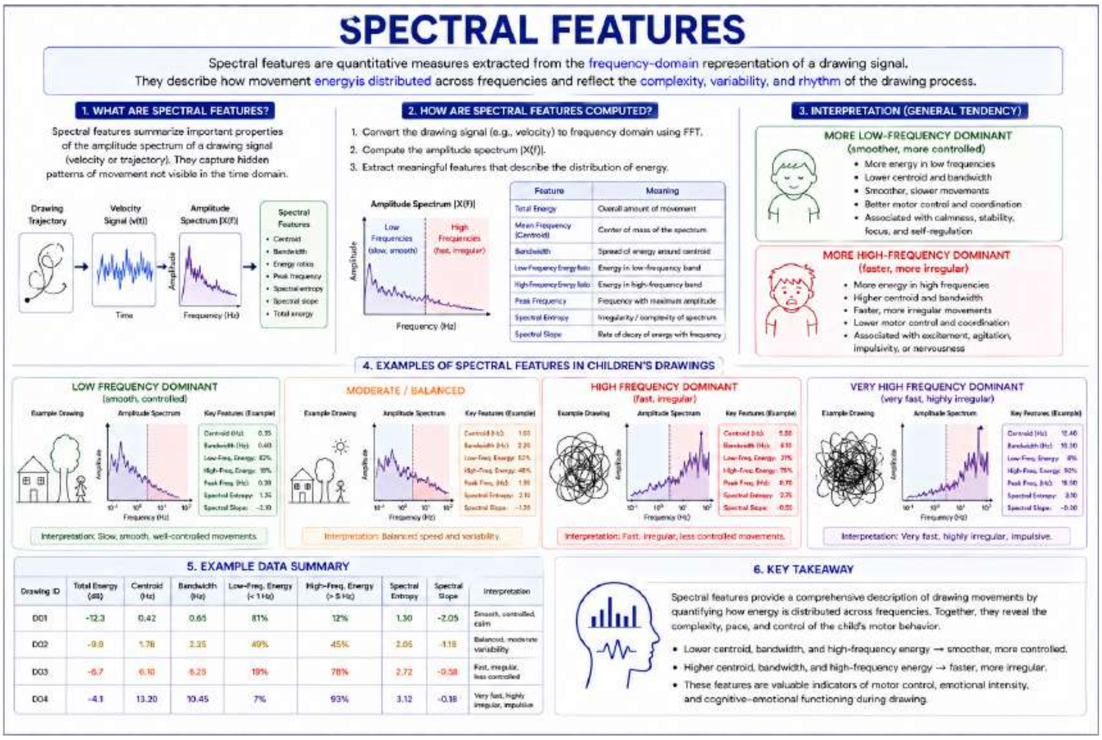


Figure 3.24: Spectral Features

3.3.7 Geometric Features

3.3.7.1 Line Inclination

Line Inclination measures the average orientation of the drawing strokes relative to the horizontal axis. It indicates whether the drawing is composed mainly of horizontal, vertical, or diagonal movements. This feature describes the overall direction of the child's drawing strokes and helps characterize the geometric structure of the drawing.

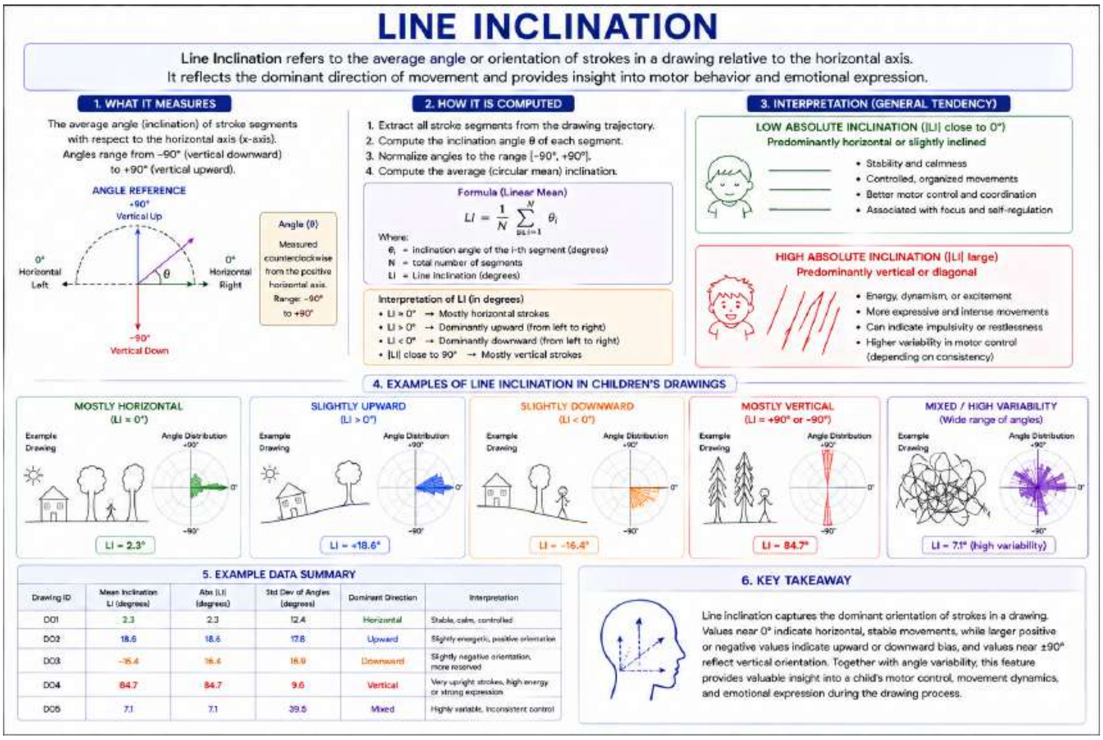


Figure 3.25: Line Inclination Feature

3.3.7.2 Curvature

Curvature measures how much a drawing stroke bends as it moves across the surface. Higher curvature values indicate more curved or irregular strokes, while lower values correspond to straighter and smoother movements. This feature describes the shape and complexity of the drawing trajectory, providing information about the child's drawing behavior.

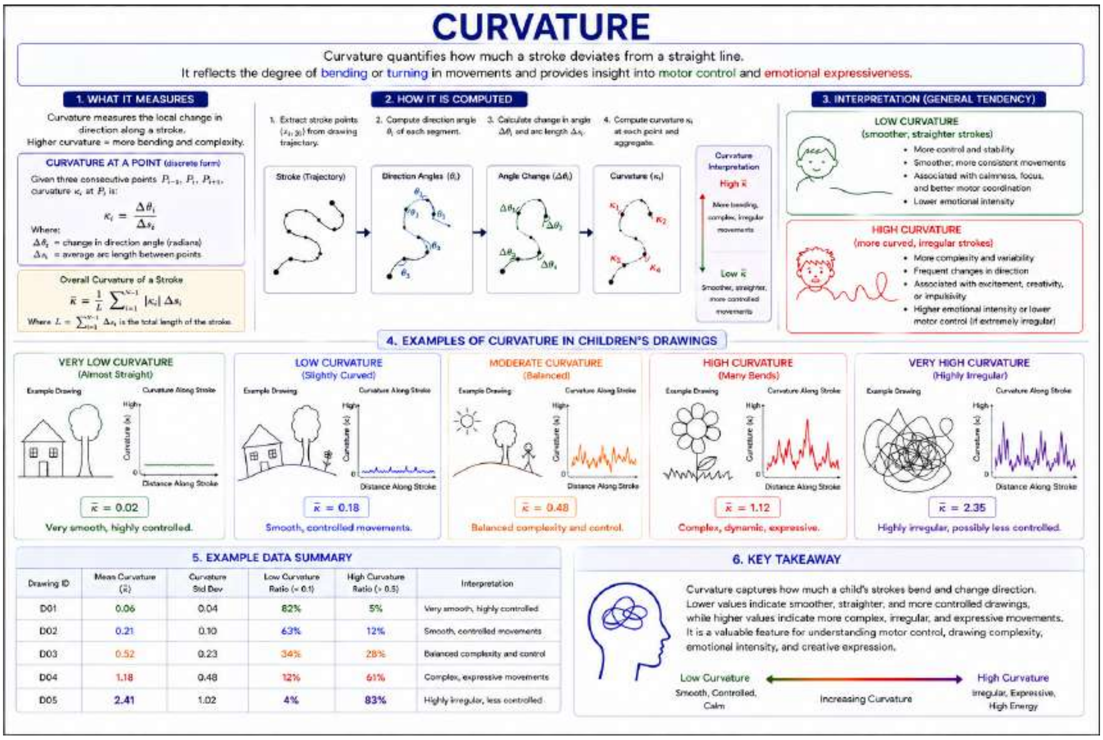


Figure 3.26: Curvature Feature

3.3.7.3 Bounding Box Area

Bounding Box Area measures the area of the smallest rectangle that completely encloses a drawing. It indicates how much of the drawing surface is occupied by the trajectory. Larger values correspond to drawings that cover a wider area, while smaller values indicate more compact drawings. This feature describes the overall spatial extent of the child's drawing.

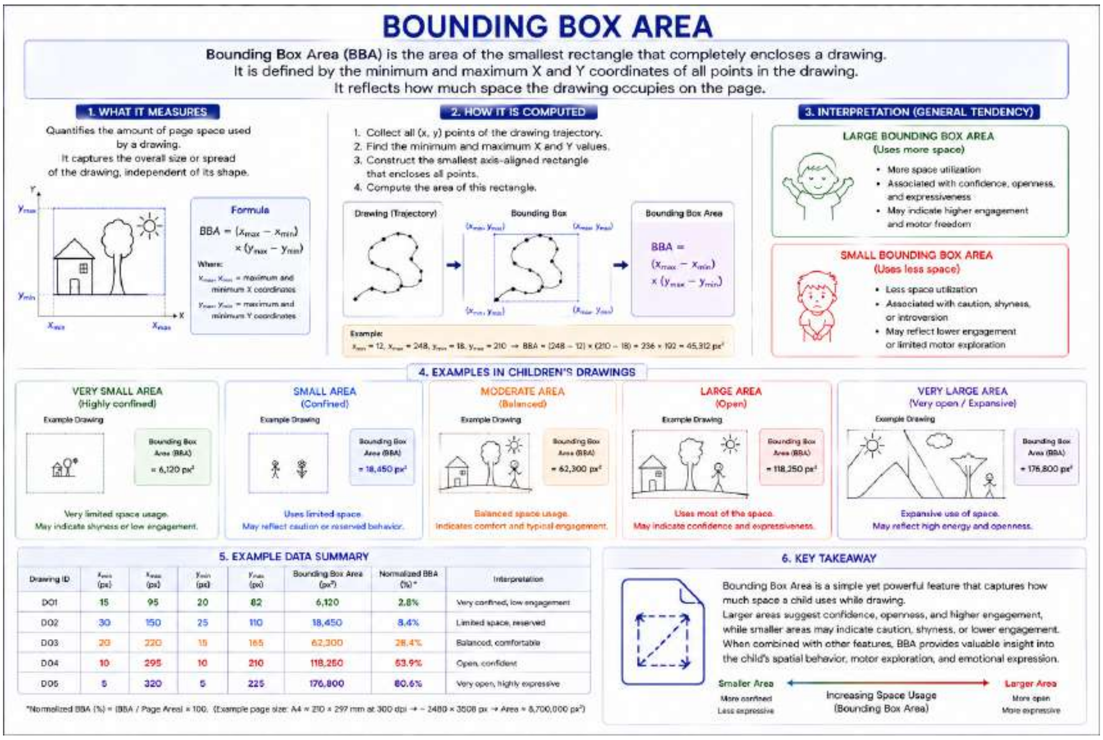


Figure 3.27: Bounding Box Area Feature

3.3.7.4 Rectangularity

Rectangularity measures how efficiently a drawing fills its bounding box. It is computed as the ratio between the drawing area (or convex hull area) and the area of the bounding box. Higher values indicate that the drawing occupies most of the bounding box, while lower values correspond to more irregular or compact shapes. This feature describes the overall spatial organization of the drawing.

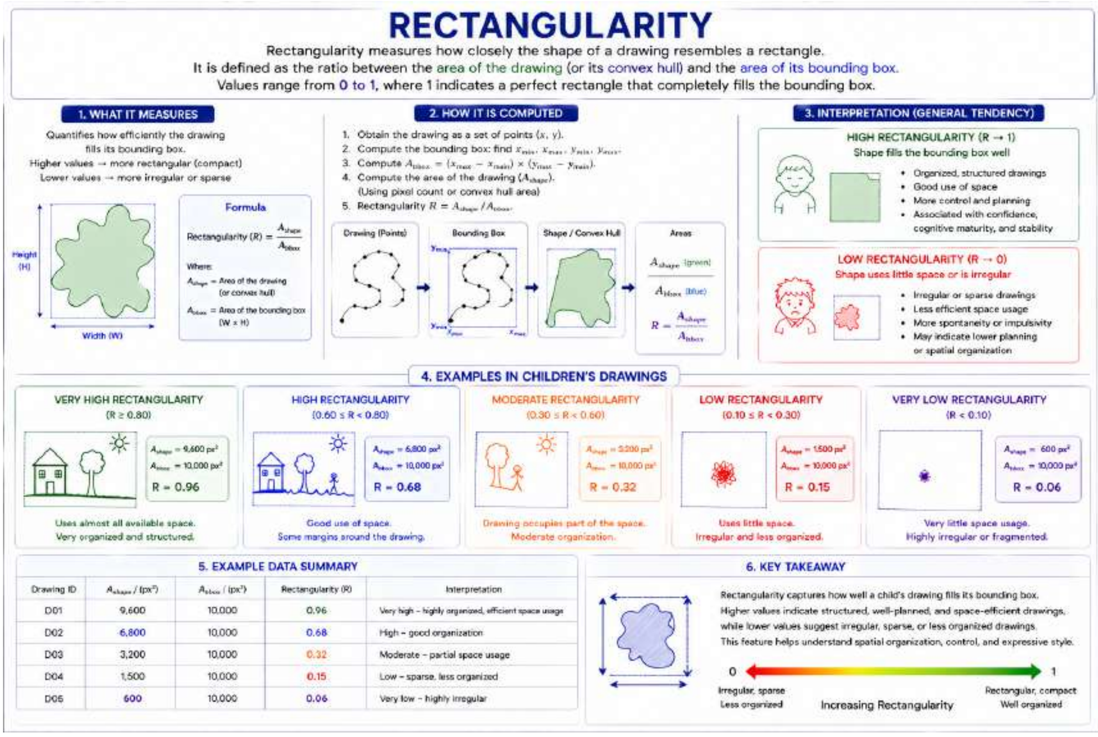


Figure 3.28: Rectangularity Feature

3.3.7.5 Compactness

Compactness measures how closely a drawing is packed by relating its area to its perimeter. Higher values indicate more compact and regular shapes, while lower values correspond to more elongated or irregular drawings. This feature describes the overall shape and spatial organization of the drawing.

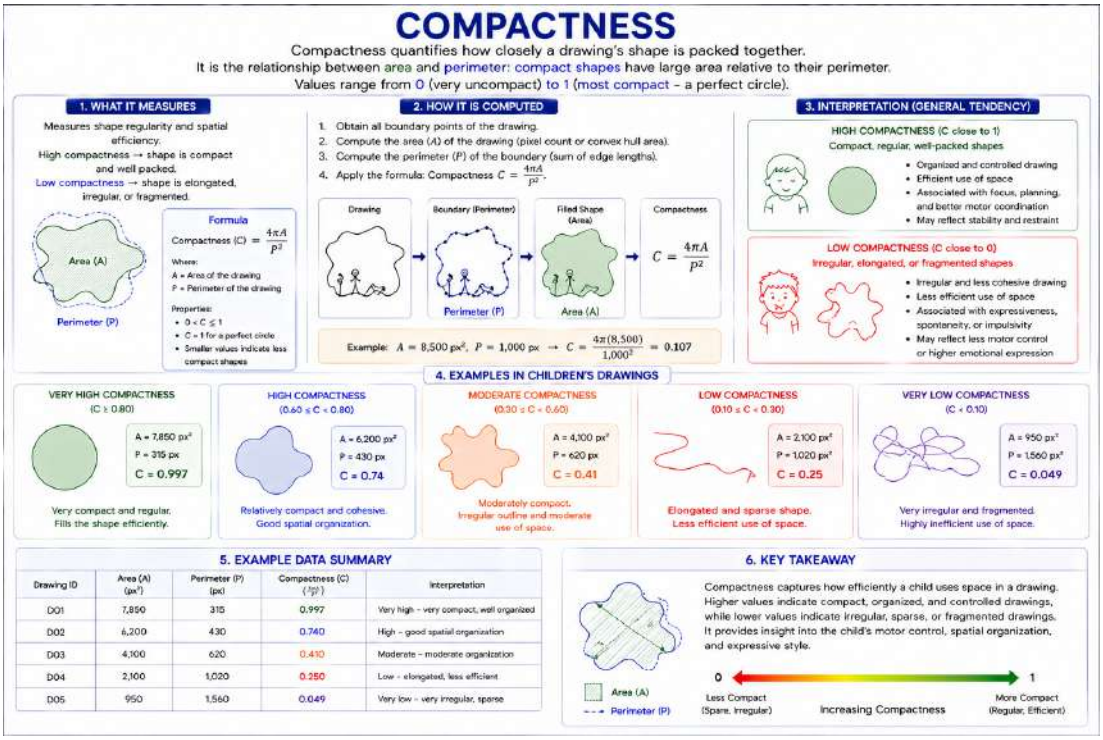


Figure 3.29: Compactness Feature

3.3.7.6 Eccentricity

Eccentricity measures how much a drawing shape differs from a perfect circle by evaluating its elongation. Values close to 0 indicate more rounded shapes, while values close to 1 correspond to more elongated shapes. This feature describes the overall shape and directional tendency of the drawing.

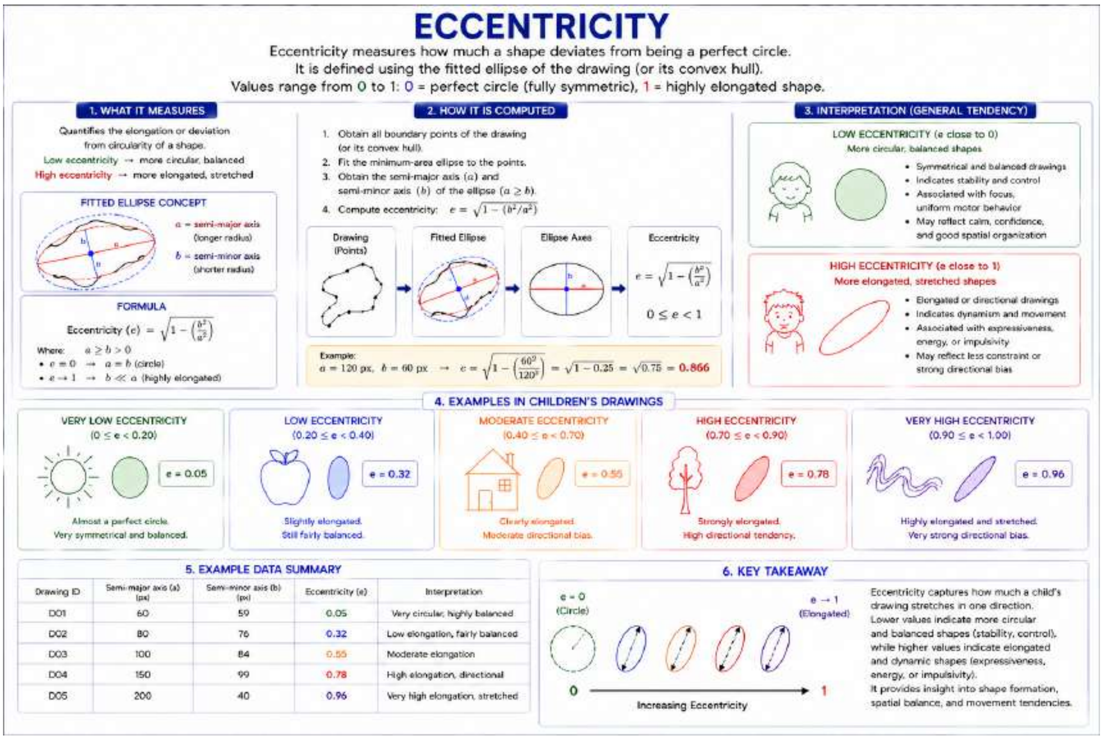


Figure 3.30: Eccentricity Feature

3.3.7.7 Convex Hull Area

Convex Hull Area measures the area of the smallest convex shape that encloses all the points in a drawing. Unlike the bounding box, it follows the outer shape of the drawing more closely. Higher values indicate that the drawing covers a larger area, while lower values correspond to more compact drawings. This feature describes the overall spatial extent of the drawing.0

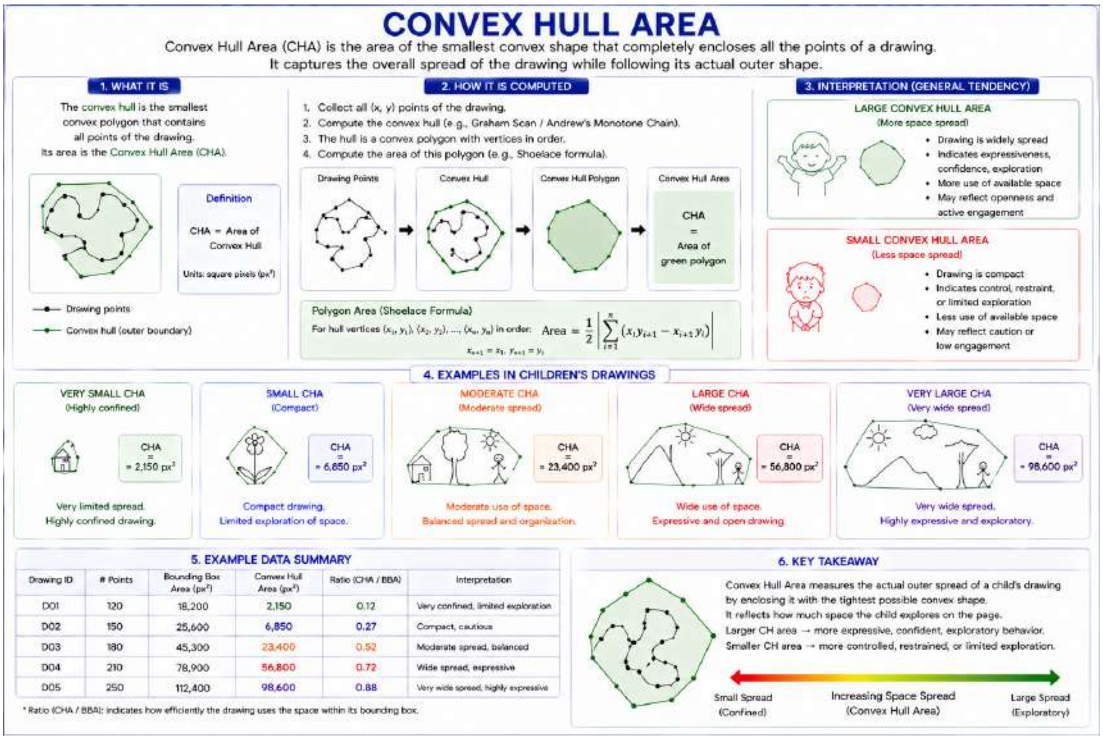


Figure 3.31: Convex Hull Area Feature

3.3.7.8 Centroid Offset

Centroid Offset measures the distance between the center of the drawing and the center of the drawing area. It indicates how far the drawing is shifted from the middle of the available space. Higher values correspond to drawings positioned farther from the center, while lower values indicate a more centered drawing. This feature describes the overall spatial placement of the drawing.

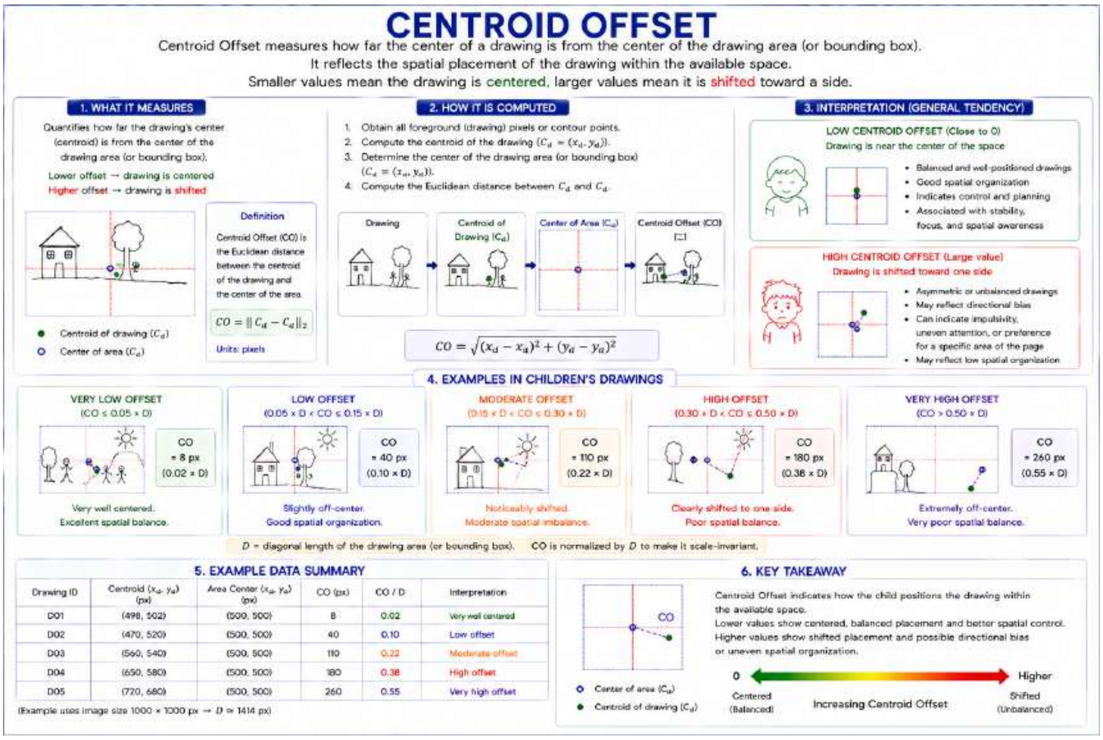


Figure 3.32: Centroid Offset Feature

3.3.7.9 Deviation from Centroid

Deviation from Centroid measures the average distance of all drawing points from the drawing's centroid. It indicates how widely the points are distributed around the center of the drawing. Higher values correspond to more spread-out drawings, while lower values indicate more compact drawings. This feature describes the overall spatial dispersion of the drawing.

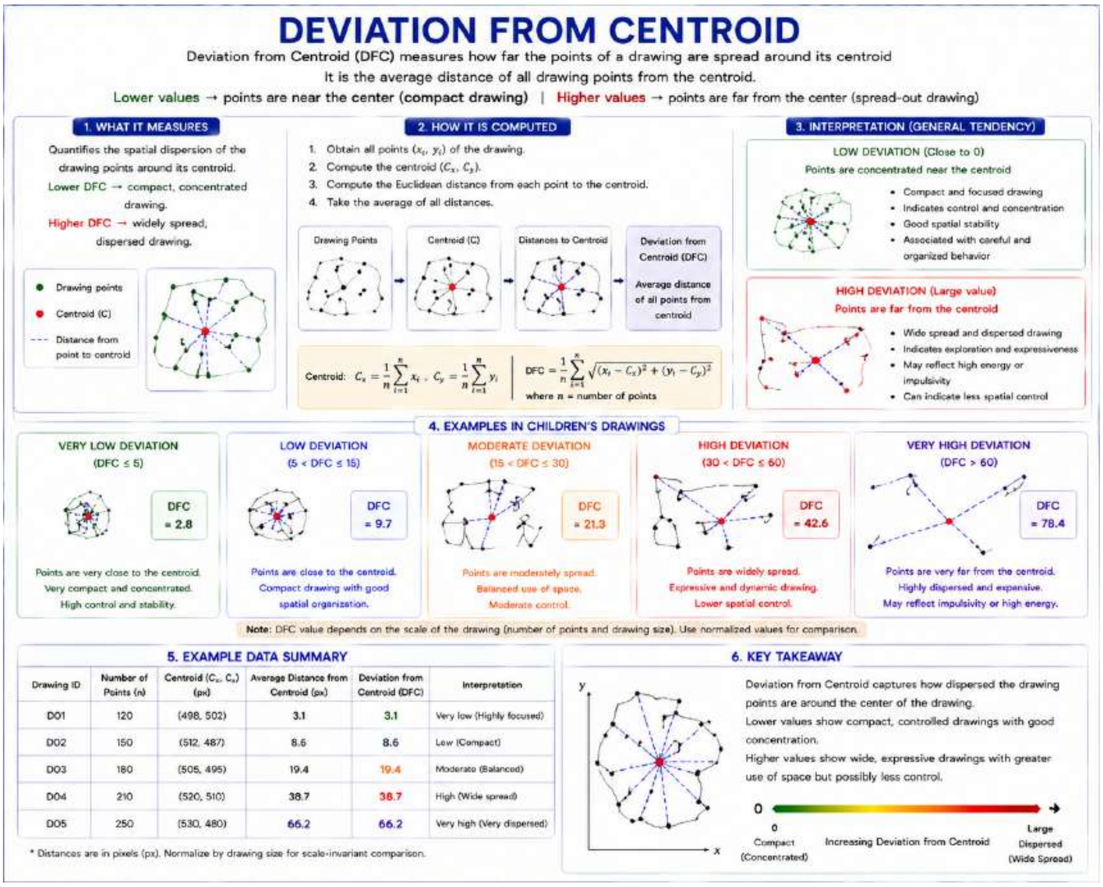


Figure 3.33: Deviation from Centroid Feature

3.3.7.10 Average Centroid Radius

Average Centroid Radius measures the average distance of all drawing points from the drawing's centroid. It indicates how far the drawing extends from its center. Higher values correspond to more expanded drawings, while lower values indicate that the drawing is concentrated around the centroid. This feature describes the overall spatial spread of the drawing.

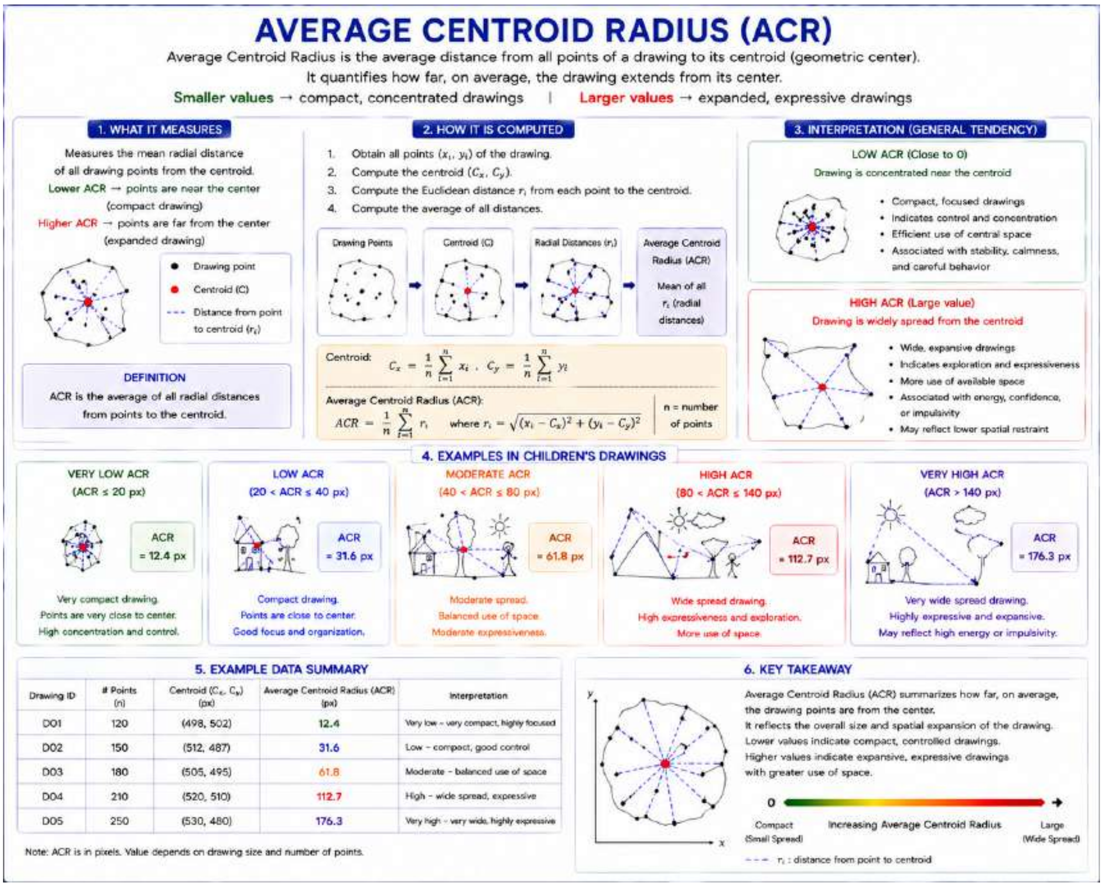


Figure 3.34: Average Centroid Radius Feature

Chapter 4

Methodology

Summary

This chapter described the methodology used to develop and evaluate the proposed emotion recognition system. It presented the dataset, preprocessing steps, feature extraction process, and the image preparation procedure. The chapter also explained the implementation of the Random Forest, MobileNetV2, and the proposed hybrid model, together with the training strategy based on nested Leave-One-Out Cross-Validation and hyperparameter optimization. Finally, the evaluation metrics used to assess model performance were introduced, providing the foundation for the experimental results presented in the next chapter.

4.1 Random Forest Approach

4.1.1 Overview of the Random Forest Regressor Model

Figure 4.1 illustrates the workflow of the Random Forest Regressor used in this thesis. The model receives the extracted drawing features as input and generates multiple decision trees using random subsets of the training data and features.

Each tree produces its own prediction, and the final output is obtained by averaging the predictions from all trees.

Unlike classification tasks, the Random Forest Regressor predicts continuous values. In this work, it simultaneously estimates the two emotional dimensions of each drawing: valence, which represents how positive or negative the emotion is, and arousal, which indicates the level of emotional activation or intensity.

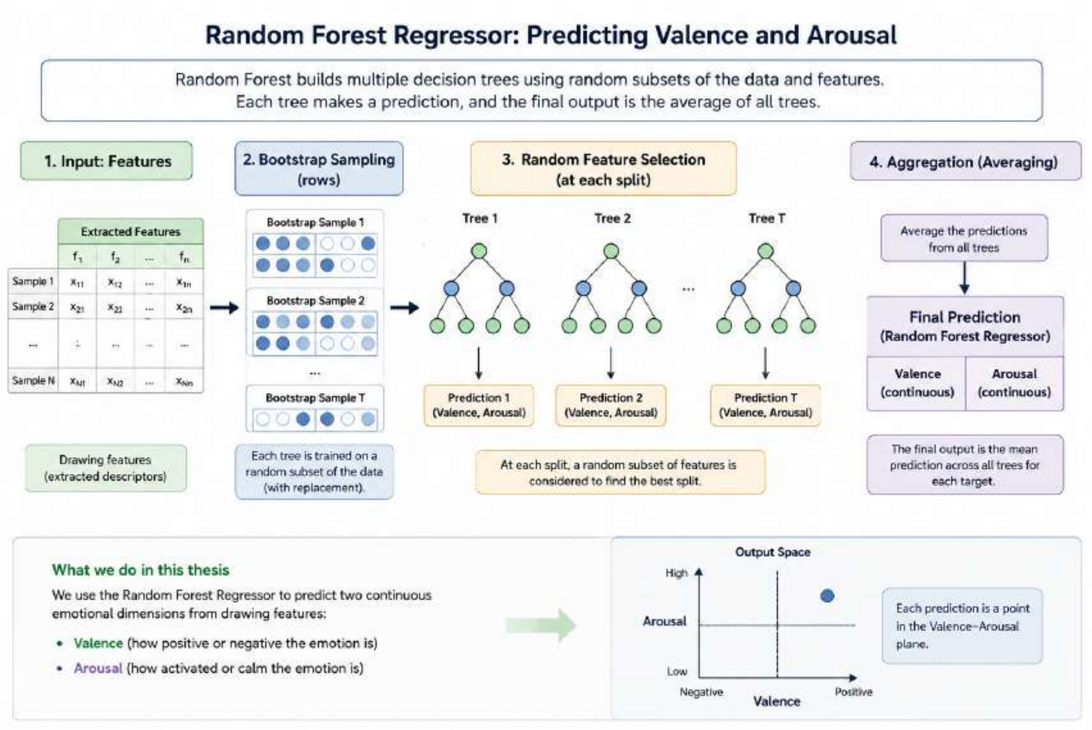


Figure 4.1: Random Forest Regressor Model

4.1.2 Valence–Arousal Emotion Grid

Figure 4.2 illustrates the two-dimensional valence–arousal model used to represent emotions. The horizontal axis represents **valence**, ranging from negative emotions on the left to positive emotions on the right. The vertical axis represents **arousal**, ranging from low emotional activation at the bottom to high emotional activation at the top.

Each emoji corresponds to a different combination of valence and arousal, showing how emotions vary in both positivity and intensity. In this thesis, the Random Forest Regressor predicts these two continuous values—**valence** and **arousal**—from the features extracted from children's drawings.

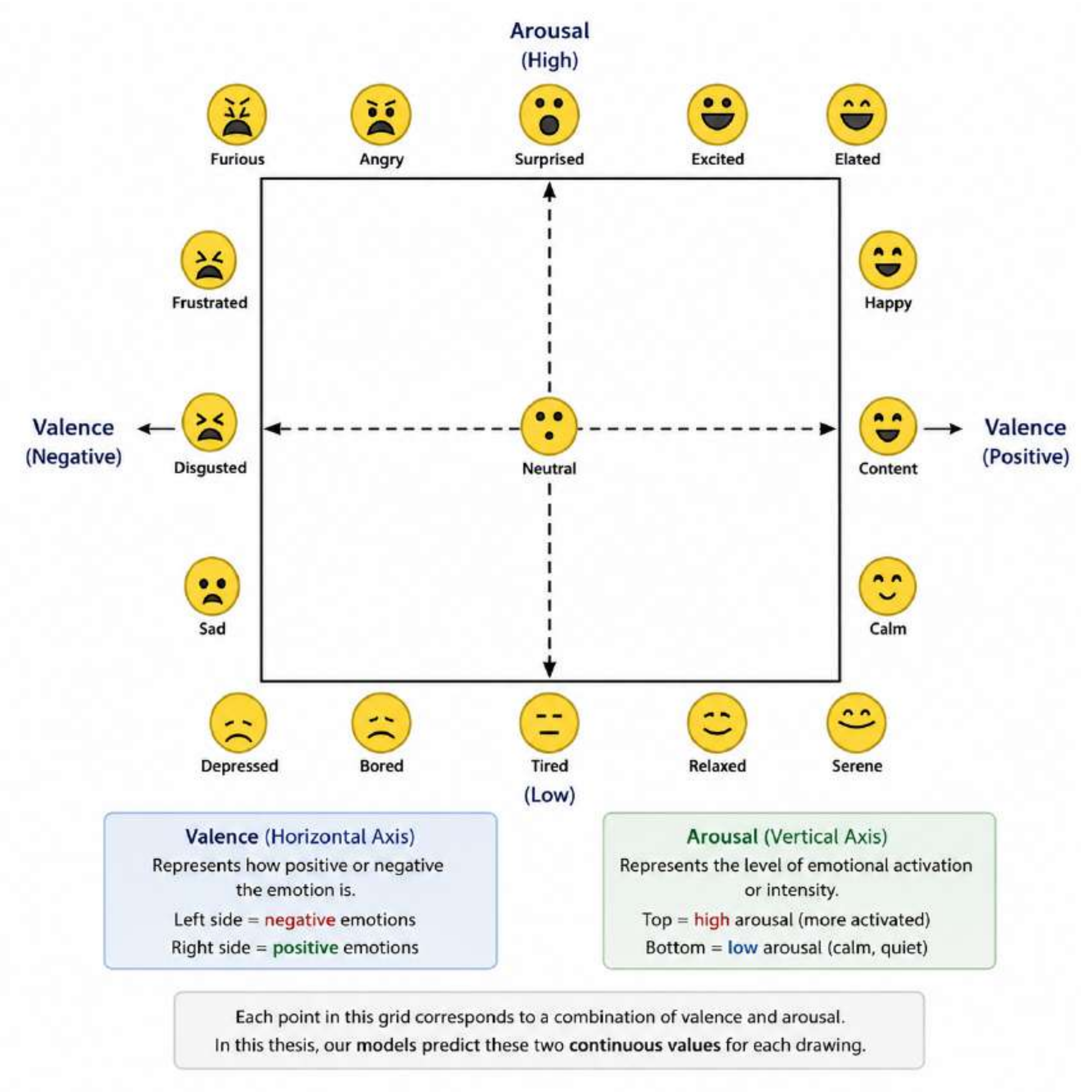


Figure 4.2: Valence–Arousal Emotion Grid

4.1.3 Dataset Creation Pipeline

The dataset was created by collecting, organizing, and matching the drawing trajectories with their corresponding emotional annotations from the TRATTO system. This produced a single structured dataset ready for machine learning.

The preprocessing pipeline was implemented in Python using the Pandas and Pathlib libraries and consisted of two main stages:

1. Annotation aggregation and normalization
2. Feature extraction and dataset construction

4.1.3.1 Annotation Aggregation and Normalization

```
def list_annotation_files(directory: str) -> list:
    p = Path(directory)
    return [str(file) for file in p.rglob('annotation*.tsv')
            if file.is_file() and "Can't_Use" not in str(file)]

annotation_files = list_annotation_files('Non_Acted_Data')

annotations = pd.DataFrame()
for file in annotation_files:
    annotations = pd.concat(
        [annotations, pd.read_csv(file, sep='\t')],
        ignore_index=True
    )
annotations.to_csv('annotations.csv', index=False)
```

Table 4.1: Generation of the annotations dataset from raw annotation files.

The annotation files were merged into a single dataframe, and the trajectory filenames were extracted and standardized to a consistent format. This ensured that each trajectory could be correctly matched with its corresponding emotional annotation.

The final annotation table contained only the trajectory filename together with the normalized valence (x) and arousal (y) coordinates and was saved as **annotations.csv**.

```
dataset_creation.ipynb    annotations.csv ×    dataset.csv
Random_Forest_Project > annotations.csv > data
trajectory_filename,y_normalized,x_normalized
25_10_18_trajectories_g_6_id_0.txt,0.4876815676689148,0.500245988368988
25_10_18_trajectories_g_5_id_0.txt,0.9023969769477844,0.8169243335723877
25_10_18_trajectories_g_5_id_1.txt,0.9128140807151794,0.8866170048713684
25_10_18_trajectories_g_5_id_2.txt,0.988441288471222,0.9481798410415648
25_10_18_trajectories_g_5_id_3.txt,0.7735217809677124,0.8896390199661255
25_10_18_trajectories_g_5_id_4.txt,0.7800723314285278,0.8053081035614014
25_10_18_trajectories_g_5_id_5.txt,0.783091127872467,0.7108943462371826
25_10_18_trajectories_g_5_id_6.txt,0.6806787252426147,0.7246490120887756
25_10_18_trajectories_g_8_id_0.txt,0.963520348072052,0.9183644652366638
```

Figure 4.3: A portion of the generated annotation table

4.1.3.2 Feature extraction and dataset construction

After creating the annotation table, the trajectory files were matched with their corresponding emotional labels to build the final machine learning dataset.

All valid ***trajectories*.txt** files were recursively collected from the **Non Acted Data** directory. For each file, the trajectory data were loaded, the **Features** module extracted the drawing descriptors, and the corresponding normalized valence and arousal labels were retrieved from the annotation table. The extracted features and labels were then combined into a single dataset entry.

Each row of the final dataset contains the extracted drawing features together with the normalized valence (*label* x) and arousal (*label* y) values.

```
annotations = pd.read_csv("annotations.csv")
features_processor = Features()
dataset = pd.DataFrame()

for trajectory_file in trajectory_files:

    trajectory_df = pd.read_csv(trajectory_file, sep="\t")
    features_row = features_processor.compute_features(trajectory_df)

    label = annotations.loc[
        annotations["trajectory_filename"] == Path(trajectory_file).name,
        ["x_normalized", "y_normalized"]
    ].iloc[0]

    features_row["label_x_normalized"] = label["x_normalized"]
    features_row["label_y_normalized"] = label["y_normalized"]

    dataset = pd.concat([dataset, features_row], ignore_index=True)

dataset.to_csv("dataset.csv", index=False)
```

Table 4.2: Generation of the final dataset through feature extraction and annotation merging.

The complete dataset was finally exported as **dataset.csv**, resulting in a dataset containing **134 samples** and **342 features** (shape: **134** × **342**).

Figure 4.4: The generated dataset structure

4.1.4 Data Preprocessing

After preprocessing, the dataset was split into training and testing sets to evaluate the performance of the Random Forest model. An 80:20 train–test split was applied using Scikit-learn, with 80% of the data used for training and the remaining 20% for testing. A fixed random seed (`SEED = 41`) was used to ensure that the split was reproducible.

The following code snippet shows the dataset splitting procedure used in this study.

```
SEED = 41
np.random.seed(SEED)

# Load data
DATA_PATH = Path('dataset.csv')
df = pd.read_csv(DATA_PATH).reset_index(drop=True)

print('Original shape:', df.shape)

dup_count = int(df.duplicated().sum())
print('Duplicate rows:', dup_count)

# Remove exact duplicate rows to reduce leakage/noise
df = df.drop_duplicates().reset_index(drop=True)

print('Shape after deduplication:', df.shape)
```

Table 4.3: Loading and preprocessing the dataset before model training.

4.1.5 Data Splitting

After preprocessing, the dataset was divided into training and testing subsets in order to evaluate the performance of the Random Forest model on unseen data.

A train-test split strategy was applied using the (train test split) function from the Scikit-learn library. In this work, 80% of the dataset was used for training, while the remaining 20% was reserved for testing. The training set was used to train the Random Forest model, whereas the testing set was used to evaluate the generalization capability of the model.

A fixed random seed (SEED = 41) was used during the splitting process to ensure reproducibility of the experiments. This guarantees that the same samples are assigned to the training and testing subsets across multiple executions.

The following code snippet shows the dataset splitting procedure used in this study:

```
test_size = 0.2  # 20% for test set

train_df, test_df = train_test_split(
    df,
    test_size=test_size,
    random_state=SEED
)

test_df
```

Table 4.4: Splitting the dataset into training and test subsets.

4.1.6 Multi-Output Random Forest Regressor

To predict the emotional state associated with each drawing, a Multi-Output Random Forest Regressor was used to simultaneously estimate the two target variables: *valence* (`label_x_normalized`) and *arousal* (`label_y_normalized`). Since both targets are continuous values, the task was formulated as a multi-output regression problem.

The model was trained using the extracted drawing features described in Chapter 3. Before training, the target columns were removed from the input data, only numerical features were retained, and missing values were replaced using `np.nan_to_num()`.

The Random Forest model was implemented using Scikit-learn's RandomForestRegressor with `n_estimators = 500`, `random_state = SEED`, `n_jobs = -1`, and `oob_score = True`. These settings improve prediction stability, ensure reproducibility, enable parallel processing, and provide an out-of-bag estimate of the model's performance.

After training, the model predicted a valence and arousal value for each sample in the test set. These predictions were then compared with the ground-truth annotations to evaluate the model's ability to estimate emotions from children's drawing features.

```
target_cols = ["label_x_normalized", "label_y_normalized"]
X_train = train_df.drop(columns=target_cols).select_dtypes(include=[np.number])
y_train = train_df[target_cols].values

X_test = test_df[X_train.columns]
y_test = test_df[target_cols].values

X_train = np.nan_to_num(X_train.values)
X_test = np.nan_to_num(X_test.values)

rf = RandomForestRegressor(
    n_estimators=500,
    random_state=SEED,
    n_jobs=-1,
    oob_score=True
)
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)
```

Table 4.5: Training and evaluating the Multi-Output Random Forest Regressor on the training and test datasets.

4.1.7 Hyperparameter Tuning

To improve the performance of the baseline Random Forest model, hyperparameter tuning was performed using `GridSearchCV`. Different combinations of key parameters, including the number of trees, tree depth, minimum samples for splitting and leaf nodes, and the number of features considered at each split, were evaluated to identify the best-performing configuration.

The optimization used Leave-One-Out Cross-Validation (LOOCV), where one sample is used for validation and the remaining samples for training. This approach is well suited for small datasets, as it allows the model to learn from nearly all available data during each iteration.

The best parameter combination was selected based on the lowest Root Mean Squared Error (RMSE). The optimized model was then evaluated on the test set and used to predict the valence and arousal values of each drawing.

The following code snippet shows the hyperparameter tuning procedure used in this study.

```
param_grid = {
    "n_estimators": [100, 250, 300],
    "max_depth": [1, 5, 10],
    "min_samples_split": [2, 5, 10],
    "min_samples_leaf": [1, 2, 3],
    "max_features": ["log2"]
}

grid = GridSearchCV(
    estimator=RandomForestRegressor(random_state=SEED),
    param_grid=param_grid,
    cv=LeaveOneOut(),
    scoring="neg_root_mean_squared_error"
)

grid.fit(X_train, y_train)

best_model = grid.best_estimator_
```

Table 4.6: Hyperparameter tuning of the Random Forest Regressor using Grid Search and Leave-One-Out Cross-Validation.

4.1.8 Model evaluation metrics

4.1.8.1 Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) measures the average absolute difference between the predicted and actual values. It provides a simple and intuitive measure of prediction accuracy, where lower values indicate better performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4.1}$$

where:

- n is the number of samples,
- y_i is the true value,
- $\hat{y}_i$ is the predicted value.

In this study, MAE was computed both as an overall score and separately for the valence and arousal predictions.

```
overall_mae = mean_absolute_error(y_test, y_pred)

mae_per_target = mean_absolute_error(
    y_test,
    y_pred,
    multioutput="raw_values"
)
```

Table 4.7: Computation of the Mean Absolute Error (MAE) for the overall model and for each target variable.

4.1.8.2 Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) measures the average difference between the predicted and actual values, giving greater weight to larger errors. Lower RMSE values indicate better model performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{4.2}$$

where:

- n is the number of samples,
- y_i is the true value,
- $\hat{y}_i$ is the predicted value.

In this study, RMSE was calculated both as an overall score and separately for the valence and arousal predictions.

```
overall_rmse = np.sqrt(
    mean_squared_error(y_test, y_pred)
)

rmse_per_target = np.sqrt(
    mean_squared_error(
        y_test,
        y_pred,
        multioutput="raw_values"
    )
)
```

Table 4.8: Computation of the Root Mean Squared Error (RMSE) for the overall model and for each target variable.

4.1.8.3 Coefficient of Determination (R²)

The Coefficient of Determination (R^2) measures how well the model explains the variation in the target values. Higher R^2 values indicate better predictive performance, with $R^2 = 1$ representing a perfect fit and $R^2 = 0$ indicating that the model performs no better than predicting the mean.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.3}$$

where:

- y_i is the true value,
- $\hat{y}_i$ is the predicted value,

- $\bar{y}$ is the mean of the true values,
- n is the number of samples.

In this study, the R^2 score was calculated both as an overall score and separately for the valence and arousal predictions.

```
overall_r2 = r2_score(y_test, y_pred)

r2_per_target = r2_score(
    y_test,
    y_pred,
    multioutput="raw_values"
)
```

Table 4.9: Computation of the coefficient of determination (R^2) for the overall model and for each target variable.

4.1.9 Leave-One-Out Cross-Validation (LOOCV)

After preparing the dataset, Leave-One-Out Cross-Validation (LOOCV) was used to evaluate the generalization performance of the Random Forest Regressor. The dataset was loaded from `dataset.csv`, duplicate samples were removed, and the drawing features were used as inputs to predict the two target variables: valence (`label_x_normalized`) and arousal (`label_y_normalized`).

In LOOCV, one sample is used for testing while all remaining samples are used for training. This process is repeated until every sample has been used once as the test sample. For each iteration, a Random Forest Regressor with 500 decision trees was trained, and the predicted valence and arousal values were recorded.

The final predictions were compared with the ground-truth values using the evaluation metrics described previously (MAE, RMSE, and R^2), both overall and separately for valence and arousal.

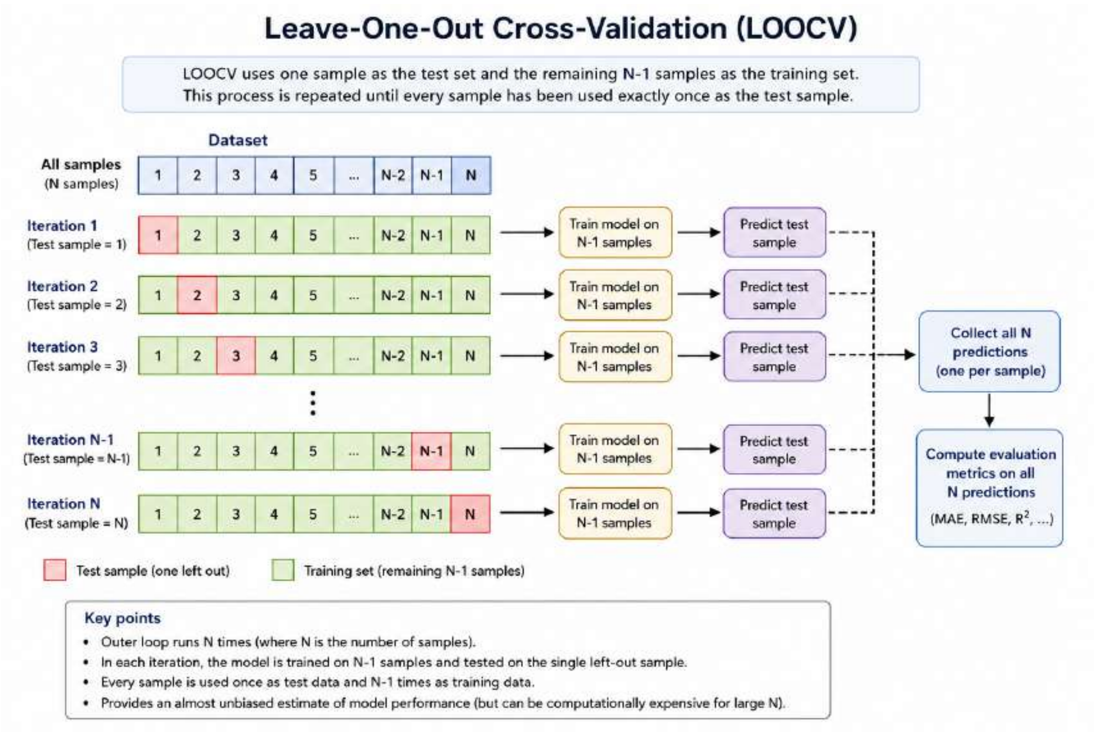


Figure 4.5: Leave-One-Out Cross-Validation Algorithm

```
target_cols = [
    "label_x_normalized",
    "label_y_normalized"]
X = df.drop(columns=target_cols).select_dtypes(include=[np.number])
X = np.nan_to_num(X.values)
y = df[target_cols].values
loo = LeaveOneOut()
y_pred = np.empty_like(y, dtype=float)
for train_idx, test_idx in loo.split(X):
    rf = RandomForestRegressor(
        n_estimators=500,
        random_state=SEED,
        n_jobs=-1)
    rf.fit(X[train_idx], y[train_idx])
    y_pred[test_idx[0]] = rf.predict(X[test_idx])[0]
overall_mae = mean_absolute_error(y, y_pred)
overall_rmse = np.sqrt(mean_squared_error(y, y_pred))
overall_r2 = r2_score(y, y_pred)
```

Table 4.10: Leave-One-Out Cross-Validation (LOOCV) Implementation for Random Forest Regression

4.1.10 Nested LOOCV with an inner GridSearchCV for Hyperparameter Tuning

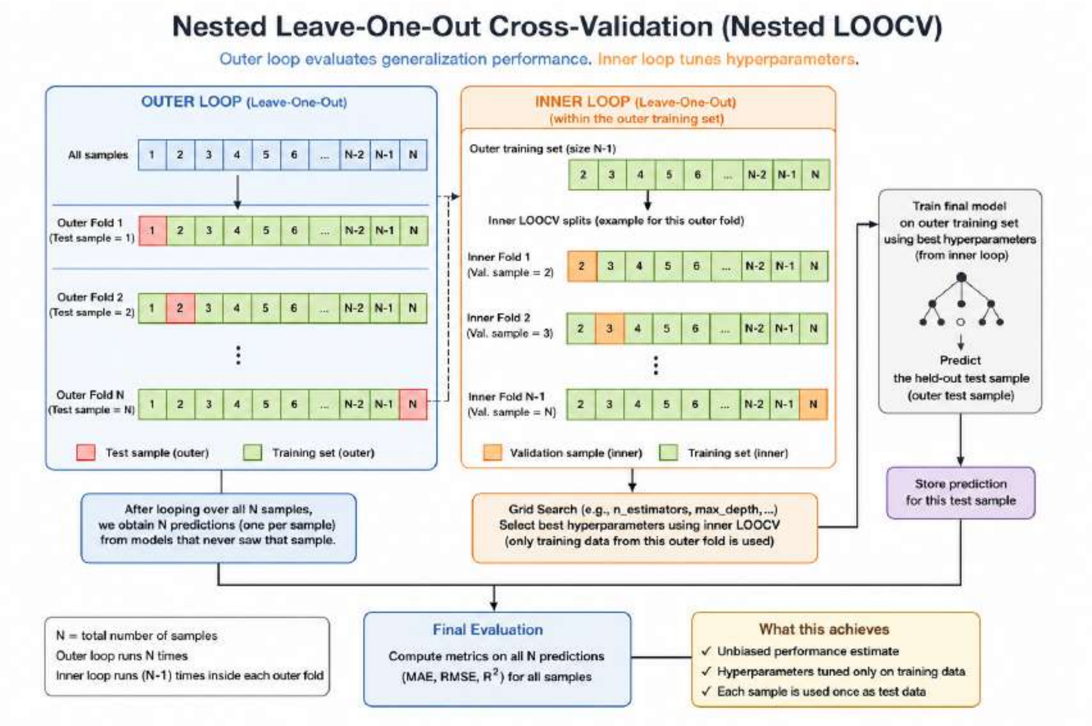


Figure 4.6: Nested Leave-One-Out Cross-Validation Algorithm

To obtain a more reliable estimate of the model's performance, a nested Leave-One-Out Cross-Validation (LOOCV) approach was used. The outer LOOCV loop evaluated the model, while the inner LOOCV loop performed hyperparameter tuning using `GridSearchCV`.

During each outer iteration, one sample was held out for testing and the remaining samples were used for training. The inner loop searched for the best combination of Random Forest hyperparameters using only the training data, ensuring that the test sample was never involved in the tuning process.

The best hyperparameters were then used to train a new Random Forest model, which predicted the valence and arousal values of the held-out sample.

This process was repeated until every sample had been used once as the test sample.

The final predictions were evaluated using MAE, RMSE, and R^2, providing an unbiased estimate of the model's ability to generalize to unseen data.

```
outer_loo = LeaveOneOut()
inner_loo = LeaveOneOut()

param_grid = {
    "n_estimators": [100, 250, 300],
    "max_depth": [1, 5, 10],
    "min_samples_split": [2, 5],
    "min_samples_leaf": [1, 2],
    "max_features": ["log2"],
}

y_pred = np.empty_like(y, dtype=float)

for train_idx, test_idx in outer_loo.split(X):

    search = GridSearchCV(
        RandomForestRegressor(random_state=SEED),
        param_grid=param_grid,
        cv=inner_loo,
        scoring="neg_mean_absolute_error"
    )

    search.fit(X[train_idx], y[train_idx])

    best_model = RandomForestRegressor(
        random_state=SEED,
        **search.best_params_
    )

    best_model.fit(X[train_idx], y[train_idx])

    y_pred[test_idx[0]] = best_model.predict(X[test_idx])[0]
```

Table 4.11: Nested Leave-One-Out Cross-Validation with Grid Search for hyperparameter selection in the Random Forest Regressor.

4.1.11 Randomized Search with Nested LOOCV

To further optimize the Random Forest model, a nested Leave-One-Out Cross-Validation (LOOCV) approach combined with `RandomizedSearchCV` was used. The outer LOOCV loop evaluated the model, while the inner LOOCV loop searched for the best hyperparameters by randomly sampling different parameter combinations.

During each outer iteration, one sample was held out for testing and the remaining samples were used for training. The inner loop explored a broad range of Random Forest hyperparameters, including the number of trees, tree depth, minimum samples for splitting and leaf nodes, and the number of features considered at each split. The best parameter set was then used to train a new model and predict the valence and arousal values of the held-out sample.

This process was repeated until every sample had been used once as the test sample. The final predictions were evaluated using MAE, RMSE, and R^2, providing an unbiased estimate of the model's performance while exploring a wider hyperparameter space than Grid Search.

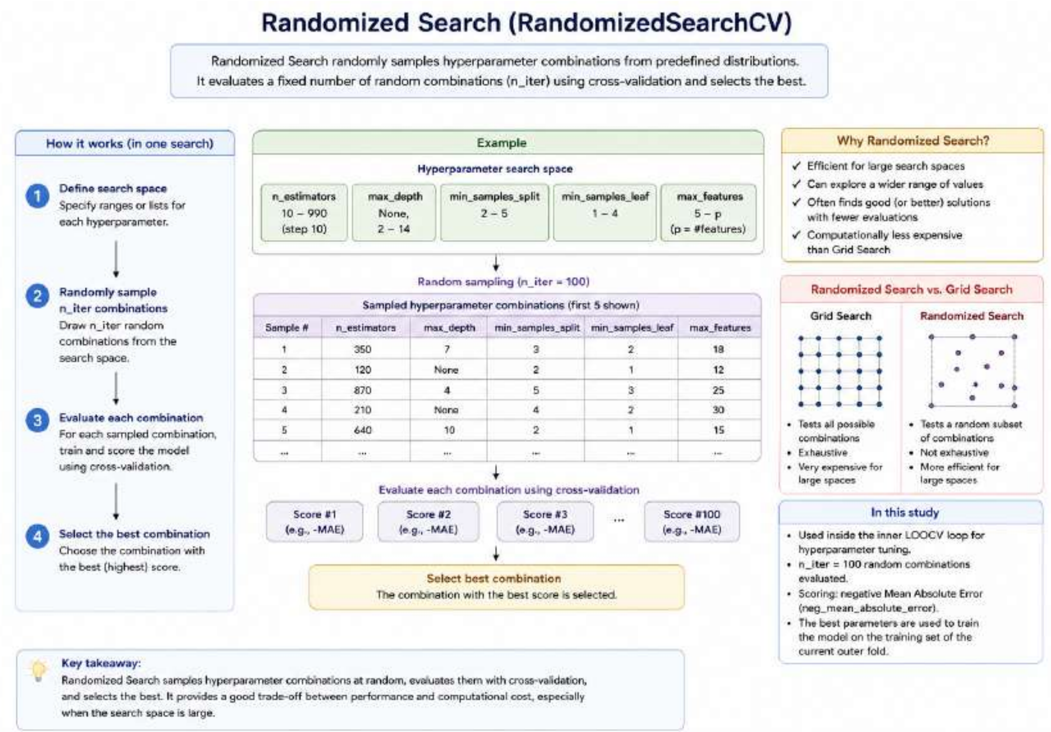


Figure 4.7: Nested LOOCV with Randomized Search

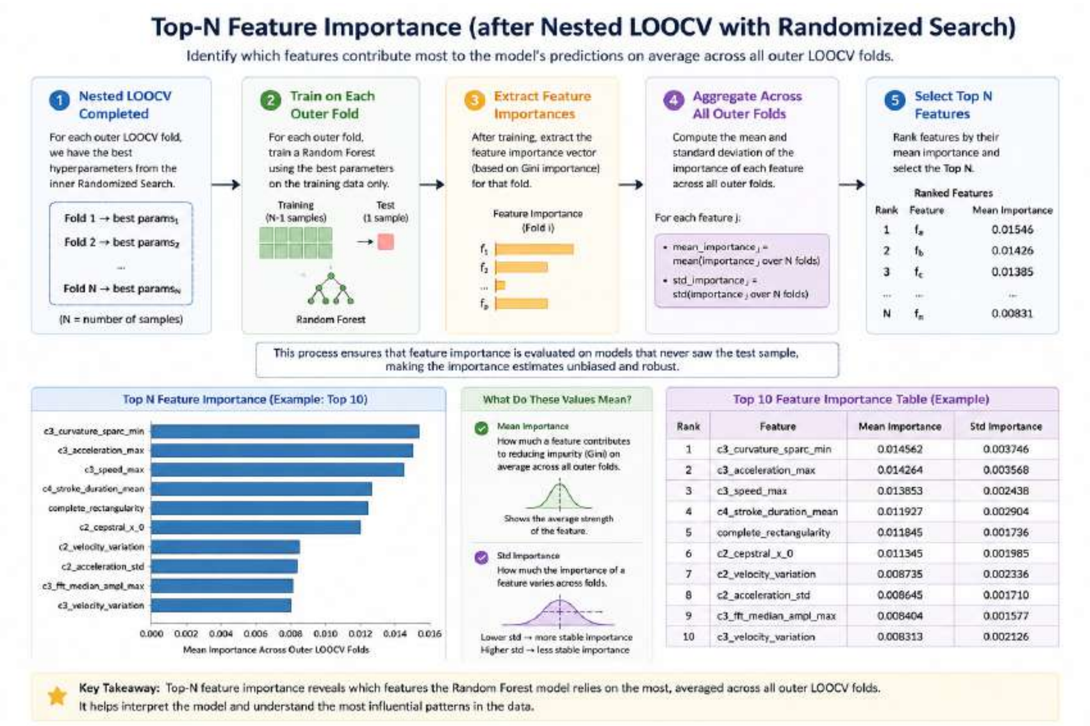
```
outer_loo = LeaveOneOut()
inner_loo = LeaveOneOut()

param_grid = {
    "n_estimators": list(range(10, 1000, 10)),
    "max_depth": [None] + list(range(2, 15)),
    "min_samples_split": list(range(2, 6)),
    "min_samples_leaf": list(range(1, 5)),
    "max_features": list(range(5, X.shape[1] + 1)),
}

y_pred = np.empty_like(y, dtype=float)

for train_idx, test_idx in outer_loo.split(X):

    search = RandomizedSearchCV(
        estimator=RandomForestRegressor(random_state=SEED),
        param_distributions=param_grid,
        n_iter=100,
        cv=inner_loo,
        scoring="neg_mean_absolute_error",
        random_state=SEED
    )

    search.fit(X[train_idx], y[train_idx])

    best_model = RandomForestRegressor(
        random_state=SEED,
        **search.best_params_
    )

    best_model.fit(X[train_idx], y[train_idx])

    y_pred[test_idx[0]] = best_model.predict(X[test_idx])[0]
```

Table 4.12: Nested Leave-One-Out Cross-Validation (LOOCV) with Randomized Search for Random Forest Hyperparameter Tuning

4.1.12 Feature Importance Analysis

After the nested LOOCV with Randomized Search, a feature importance analysis was performed to identify the drawing features that contributed most to the Random Forest predictions. For each outer fold, the best model was used to compute feature importance scores, which were then averaged across all folds. The standard deviation was also calculated to measure the consistency of each feature's importance.

The top ten features were visualized using a horizontal bar chart. The results showed that features related to drawing dynamics, such as acceleration, speed, stroke duration, and velocity variation, together with geometric features such as curvature and rectangularity, had the greatest influence on the predictions. This suggests that both the movement patterns and the structural characteristics of children's drawings provide valuable information for estimating valence and arousal.



Rank	Feature	Mean Importance	Std Importance
1	c3_curvature_sparc_min	0.014562	0.003746
2	c3_acceleration_max	0.014264	0.003568
3	c3_speed_max	0.013853	0.002438
4	c4_stroke_duration_mean	0.011927	0.002904
5	complete_rectangularity	0.011845	0.001736
6	c2_cepstral_x_0	0.011345	0.001985
7	c2_velocity_variation	0.008735	0.002336
8	c2_acceleration_std	0.008645	0.001710
9	c3_fft_median_ampl_max	0.008404	0.001577
10	c3_velocity_variation	0.008313	0.002126

Figure 4.8: Feature Importance Analysis

```
fi_per_fold = []

for fold, (train_idx, _) in enumerate(outer_loo.split(X), start=1):

    model = RandomForestRegressor(
        random_state=SEED,
        **best_params_per_fold[fold - 1]
    )

    model.fit(X[train_idx], y[train_idx])
    fi_per_fold.append(model.feature_importances_)

fi_df = pd.DataFrame({
    "feature": X_df.columns,
    "mean_importance": np.mean(fi_per_fold, axis=0),
    "std_importance": np.std(fi_per_fold, axis=0)
}).sort_values("mean_importance", ascending=False)

top_fi = fi_df.head(10)
```

Table 4.13: Random Forest Feature Importance Analysis Across LOOCV Folds

4.1.13 Random Forest Evaluation Using Top 10 Features

To investigate whether a smaller feature set could maintain the model's performance, an additional experiment was conducted using only the ten most important features identified by the feature importance analysis. These features were selected based on their average importance across the outer LOOCV folds.

A new dataset containing only the selected features was created, and Leave-One-Out Cross-Validation (LOOCV) was repeated. For each fold, the Random Forest Regressor was trained using the best hyperparameters obtained from the previous nested randomized search, with any feature-dependent parameters adjusted to match the reduced feature set.

The final predictions were evaluated using MAE, RMSE, and R^2, both overall and separately for valence and arousal. This experiment assessed whether a reduced set of highly informative features could achieve comparable predictive performance while simplifying the model.

```
TOP_N = 10

top_features = (
    fi_df.sort_values("mean_importance", ascending=False)
    ["feature"]
    .head(TOP_N)
    .tolist()
)

X_all = df.drop(columns=target_cols).select_dtypes(include=[np.number])

X_top = X_all[top_features]
X_top = np.nan_to_num(X_top.to_numpy())

y = df[target_cols].to_numpy()
```

Table 4.14: Leave-One-Out Cross-Validation using the top-ranked features.

4.2 MobileNet Model Approach

4.2.1 MobileNet Model Architecture

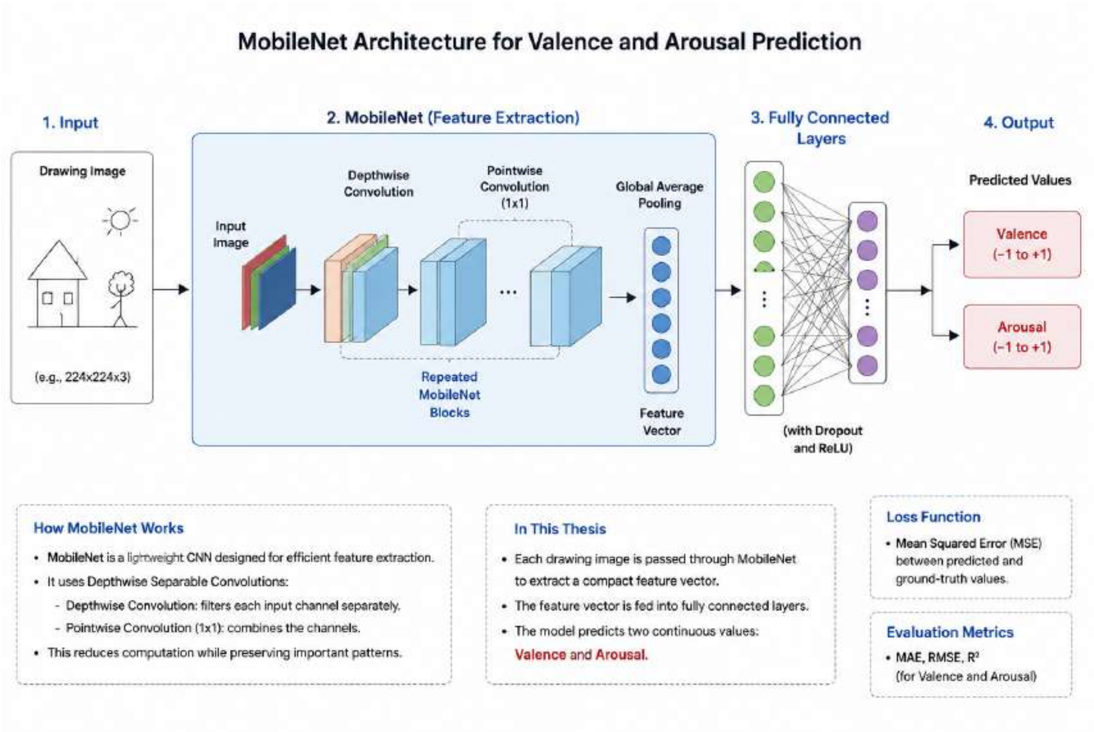
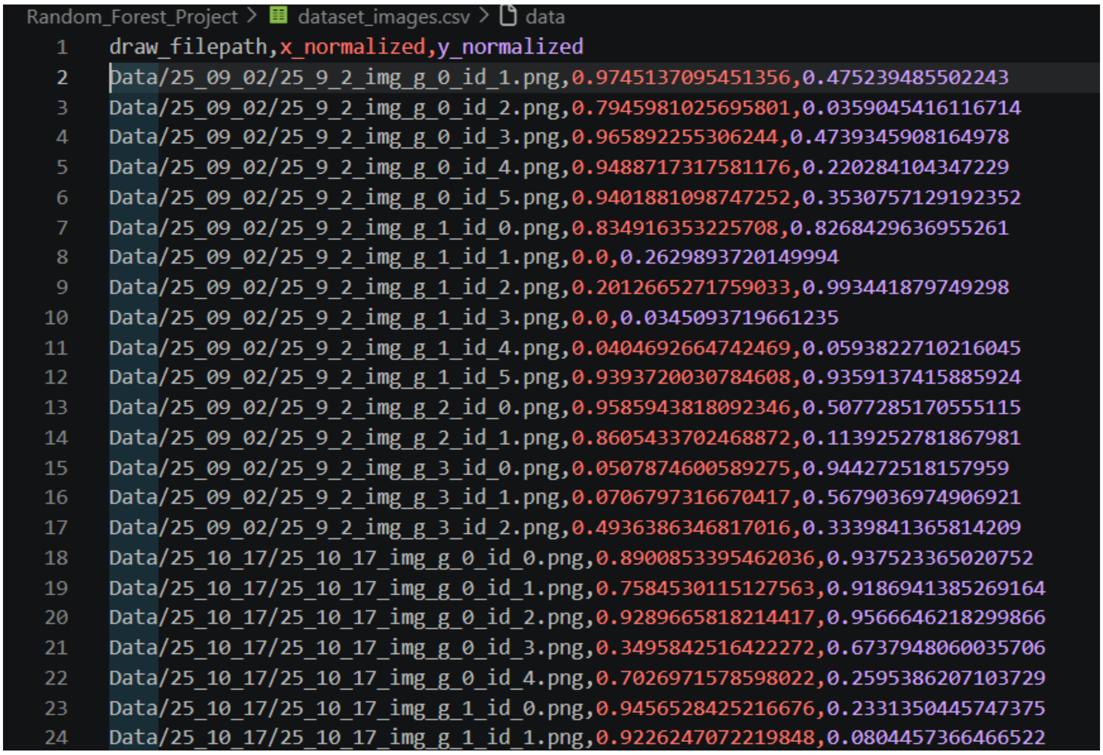


Figure 4.9: MobileNet Model Architecture

Figure 4.9 illustrates the MobileNet architecture used in this project to predict the emotional dimensions of children's drawings. Each drawing image is first passed through the MobileNet backbone, where a series of depthwise separable convolution layers extract meaningful visual features while keeping the model computationally efficient.

The extracted feature vector is then processed by fully connected layers to produce two continuous outputs: valence and arousal. These values represent the emotional position of the drawing in the valence–arousal space. The model is trained to minimize the difference between the predicted and ground-truth values, and its performance is evaluated using MAE, RMSE, and R^2.

4.2.2 Dataset Preparation for MobileNet


Random_Forest_Project > dataset_images.csv > data

```
draw_filepath,x_normalized,y_normalized
Data/25_09_02/25_9_2_img_g_0_id_1.png,0.9745137095451356,0.475239485502243
Data/25_09_02/25_9_2_img_g_0_id_2.png,0.7945981025695801,0.0359045416116714
Data/25_09_02/25_9_2_img_g_0_id_3.png,0.965892255306244,0.4739345908164978
Data/25_09_02/25_9_2_img_g_0_id_4.png,0.9488717317581176,0.220284104347229
Data/25_09_02/25_9_2_img_g_0_id_5.png,0.9401881098747252,0.3530757129192352
Data/25_09_02/25_9_2_img_g_1_id_0.png,0.834916353225708,0.8268429636955261
Data/25_09_02/25_9_2_img_g_1_id_1.png,0.0,0.2629893720149994
Data/25_09_02/25_9_2_img_g_1_id_2.png,0.2012665271759033,0.993441879749298
Data/25_09_02/25_9_2_img_g_1_id_3.png,0.0,0.0345093719661235
Data/25_09_02/25_9_2_img_g_1_id_4.png,0.0404692664742469,0.0593822710216045
Data/25_09_02/25_9_2_img_g_1_id_5.png,0.9393720030784608,0.9359137415885924
Data/25_09_02/25_9_2_img_g_2_id_0.png,0.9585943818092346,0.5077285170555115
Data/25_09_02/25_9_2_img_g_2_id_1.png,0.8605433702468872,0.1139252781867981
Data/25_09_02/25_9_2_img_g_3_id_0.png,0.0507874600589275,0.944272518157959
Data/25_09_02/25_9_2_img_g_3_id_1.png,0.0706797316670417,0.5679036974906921
Data/25_09_02/25_9_2_img_g_3_id_2.png,0.4936386346817016,0.3339841365814209
Data/25_10_17/25_10_17_img_g_0_id_0.png,0.8900853395462036,0.937523365020752
Data/25_10_17/25_10_17_img_g_0_id_1.png,0.7584530115127563,0.9186941385269164
Data/25_10_17/25_10_17_img_g_0_id_2.png,0.9289665818214417,0.9566646218299866
Data/25_10_17/25_10_17_img_g_0_id_3.png,0.3495842516422272,0.6737948060035706
Data/25_10_17/25_10_17_img_g_0_id_4.png,0.7026971578598022,0.2595386207103729
Data/25_10_17/25_10_17_img_g_1_id_0.png,0.9456528425216676,0.2331350445747375
Data/25_10_17/25_10_17_img_g_1_id_1.png,0.9226247072219848,0.0804457366466522
```

Figure 4.10: Mobile Net Dataset Generation

To prepare the dataset for the MobileNet model, all drawing images and annotation files were collected from the main data directory, while invalid folders were excluded. The annotation files provided the normalized valence (`x_normalized`) and arousal (`y_normalized`) labels associated with each drawing.

Both standard annotation files and drawing export annotation files were processed to obtain the image paths and corresponding emotional labels. Only samples with valid image files were retained, and the final dataset was saved as `dataset_images.csv`. Each row contains the image path together with its normalized valence and arousal values, which serve as the target variables for the

MobileNet model.

4.2.3 Image Loading and Preprocessing for MobileNetV2

After creating `dataset_images.csv`, the drawing images were loaded and preprocessed for the MobileNet model. A fixed random seed was used to ensure reproducibility, and each image was converted to RGB, resized to `224 × 224` pixels, and normalized to the range `[0, 1]`.

The processed images were stored in the input array `X`, while the corresponding normalized valence (`x_normalized`) and arousal (`y_normalized`) values were stored in the target array `y`. These data were then used to train the MobileNet model as a two-output regression task.

```
SEED = 42
np.random.seed(SEED)
tf.random.set_seed(SEED)

df = pd.read_csv("dataset_images.csv")

def load_and_preprocess(img_path, target_size=(224, 224)):
    img = Image.open(img_path).convert("RGB")
    img = img.resize(target_size)
    return np.array(img).astype(np.float32) / 255.0

X = np.array([load_and_preprocess(p) for p in df["draw_filepath"]])

y = df[["x_normalized", "y_normalized"]].to_numpy()
```

Table 4.15: Image Loading and Preprocessing for MobileNetV2

4.2.4 MobileNetV2 Model Training and Validation

The image dataset was split into training and validation sets using an 80:20 ratio. To improve the model's ability to generalize, light data augmentation, including small rotations, shifts, and horizontal flipping, was applied to the training images.

A pre-trained MobileNetV2 model was used as the feature extractor, and its original classification head was replaced with a regression head that predicts the normalized valence and arousal values. During the initial training phase, the MobileNetV2 layers were frozen, and only the new regression layers were trained.

The model was optimized using the Adam optimizer with Mean Squared Error (MSE) as the loss function and Mean Absolute Error (MAE) as the evaluation metric. Early stopping was used to prevent overfitting by restoring the model weights that achieved the best validation performance. After training, the model was evaluated on the validation set by comparing the predicted and true valence and arousal values.

```
X_train, X_val, y_train, y_val = train_test_split(
    X, y,
    test_size=0.2,
    random_state=SEED
)

history = model.fit(
    datagen.flow(X_train, y_train, batch_size=16),
    validation_data=(X_val, y_val),
    epochs=100,
    callbacks=[early_stop]
)

val_predictions = model.predict(X_val)

val_mae = np.mean(
    np.abs(val_predictions - y_val),
    axis=0
)
```

Table 4.16: MobileNetV2 Training and Validation

4.2.5 Final MobileNetV2 Training on the Complete Dataset

After the initial evaluation, a final MobileNetV2 model was trained using the entire dataset to make use of all available drawing samples. The pre-trained MobileNetV2 backbone was used as a feature extractor, while a custom regression head was added to predict the normalized valence and arousal values.

The model was trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function and Mean Absolute Error (MAE) as the evaluation metric. Data augmentation was applied to improve generalization, and early stopping monitored the training loss to prevent overfitting by restoring the best model weights.

```
X_full = np.concatenate([X_train, X_val], axis=0)
y_full = np.concatenate([y_train, y_val], axis=0)

base_model = tf.keras.applications.MobileNetV2(
    weights="imagenet",
    include_top=False,
    input_shape=(224, 224, 3)
)
base_model.trainable = False

x = tf.keras.layers.GlobalAveragePooling2D()(base_model.output)
x = tf.keras.layers.Dense(128, activation="relu")(x)
x = tf.keras.layers.Dropout(0.3)(x)
outputs = tf.keras.layers.Dense(2, activation="linear")(x)

model = tf.keras.Model(base_model.input, outputs)

model.compile(
    optimizer=tf.keras.optimizers.Adam(learning_rate=1e-3),
    loss="mse",
    metrics=["mae"]
)

history = model.fit(
    datagen.flow(X_full, y_full, batch_size=16),
    epochs=100,
    callbacks=[early_stop]
)
```

Table 4.17: Final MobileNetV2 Model Training on the Complete Dataset

4.3 Hybird Model Approach

4.3.1 Hybrid Model Architecture

Figure 4.11 illustrates the hybrid model proposed in this thesis. Each drawing is processed through two parallel branches. The first branch extracts the top ten handcrafted trajectory features, which are used by the Random Forest model to predict valence and arousal. The second branch processes the drawing image using MobileNetV2 to produce a second set of predictions.

The outputs of both models are then combined using a late fusion strategy based on simple averaging. The resulting fused prediction represents the final estimated valence and arousal values for each drawing, allowing the hybrid model to leverage both handcrafted features and visual information.

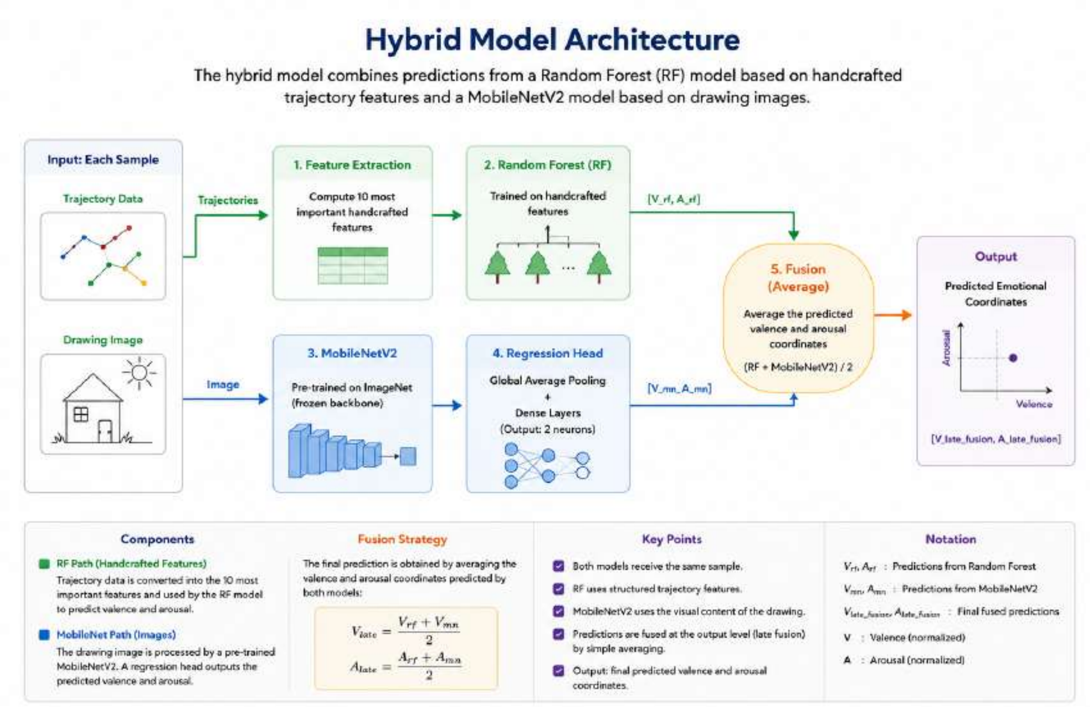


Figure 4.11: Hybrid Model Architecture

4.3.2 Hybrid Model Data Preparation

To combine the Random Forest and MobileNetV2 models, a hybrid dataset was created by matching the trajectory features with their corresponding drawing images. Only samples containing both valid handcrafted features and a matching image were included.

The Random Forest model used the top ten most important handcrafted features, while the corresponding images were resized to 224×224 pixels and normalized for MobileNetV2. Both models shared the same target variables: the normalized valence and arousal coordinates.

4.3.3 Hybrid Training Procedure

The hybrid model was evaluated using nested Leave-One-Out Cross-Validation (LOOCV). In each outer fold, one sample was used for testing while the remaining samples were used for training. Both the Random Forest and MobileNetV2 models were trained and evaluated using the same train–test split.

For the Random Forest model, an inner Grid Search with LOOCV was used to select the best hyperparameters before training the model on the full training set. At the same time, a MobileNetV2 model with ImageNet pre-trained weights was trained using the training images, with data augmentation and early stopping applied to improve generalization.

Both models then predicted the valence and arousal values of the held-out sample. This process was repeated until every sample had been tested once, producing a complete set of out-of-sample predictions. These predictions were then combined and used as the input to the final hybrid fusion model.

```
hybrid_results_df = pd.DataFrame({
    "trajectory_filename": hybrid_df["trajectory_filename"],
    "draw_filepath": hybrid_df["draw_filepath"],
    "actual_valence": y[:, 0],
    "actual_arousal": y[:, 1],
    "valence_rf": y_pred_rf[:, 0],
    "arousal_rf": y_pred_rf[:, 1],
    "valence_mobnet": y_pred_mobnet[:, 0],
    "arousal_mobnet": y_pred_mobnet[:, 1],
})
hybrid_results_df.to_csv(HYBRID_RESULTS_PATH, index=False)
```

Table 4.18: Hybird Model - Random Forest and Mobile Net

Chapter 5

Results and Plotting

Summary

This chapter presented the experimental results of the proposed emotion recognition models. It evaluated the Random Forest, MobileNetV2, and hybrid approaches, showing that the optimized Random Forest achieved the best overall performance, especially for valence prediction. Although the hybrid model improved upon MobileNetV2, it did not outperform the Random Forest. Finally, the Wilcoxon Signed-Rank Test confirmed the statistical significance of the observed performance differences, providing a solid basis for the conclusions presented in the next chapter.

5.1 Trajectory Concentration

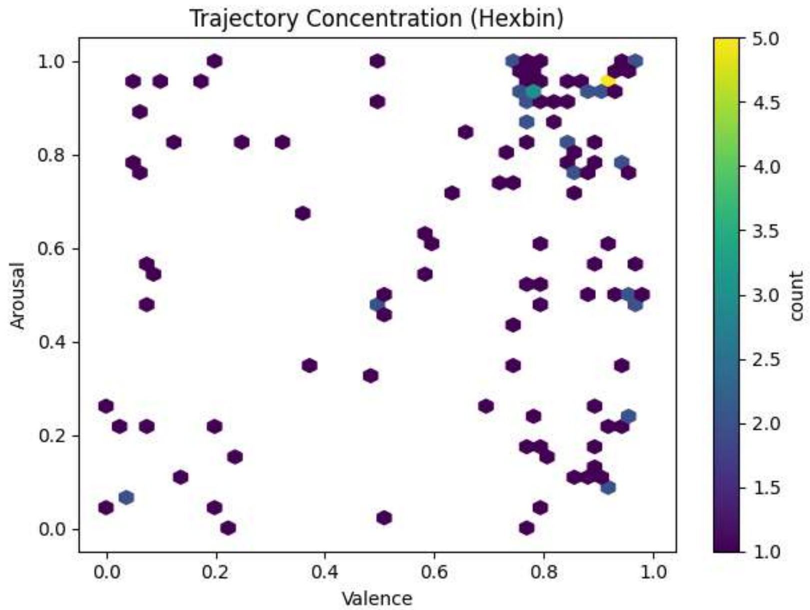


Figure 5.1: Distribution of drawings in the valence-arousal space represented using a hexagonal binning visualization.

Figure 5.1 shows the distribution of the annotated drawings in the valence-arousal space. Most samples are concentrated in the high-valence, high-arousal region, while other areas contain fewer drawings. This imbalance may lead to better predictions for well-represented emotions and lower accuracy for less frequent emotional states.

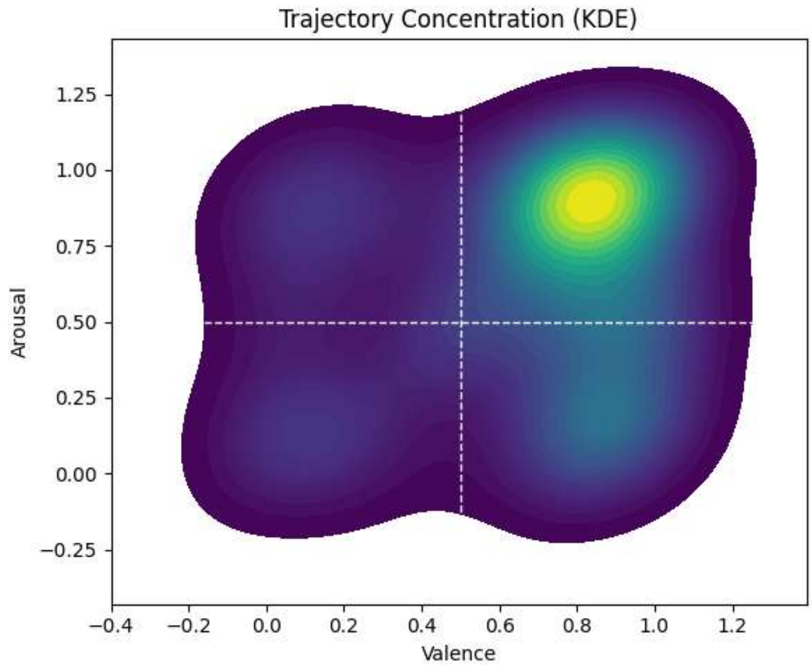


Figure 5.2: Distribution of drawings in the valence-arousal space represented using a hexagonal binning visualization.

.Figure 5.2 shows the kernel density estimation (KDE) of the drawings in the valence-arousal space. The highest concentration of samples appears in the high-valence, high-arousal region, while the remaining areas are less populated. This uneven distribution may cause the model to perform better on common emotional states and less accurately on underrepresented ones.

5.2 Random Forest Model: Leave-One-Out Cross-Validation Results

The baseline Random Forest model was evaluated using Leave-One-Out Cross-Validation (LOOCV), where each drawing was used once for testing and the remaining 129 drawings for training. The model was trained on 130 samples with

340 numerical features using 500 trees and the selected hyperparameters.

The model achieved an overall MAE of 0.243 and an RMSE of 0.296. Valence was predicted more accurately (MAE = 0.204, RMSE = 0.265, R^2 = 0.214) than arousal (MAE = 0.282, RMSE = 0.323, R^2 = -0.010). Overall, the results suggest that the extracted drawing features capture valence more effectively than arousal, while the lower performance for arousal may be related to the limited dataset and the uneven distribution of samples across the valence-arousal space.

Table 5.1: Leave-One-Out Cross-Validation performance of the Random Forest Regressor.

Target	**MAE**	**RMSE**	R^2
Valence	0.204	0.265	0.214
Arousal	0.282	0.323	-0.010
Overall	0.243	0.296	0.102

5.2.1 Nested LOOCV with an inner GridSearchCV for hyperparameter tuning

Nested Leave-One-Out Cross-Validation (Nested LOOCV) was used to obtain a more reliable estimate of the model's performance while tuning the hyperparameters within each training fold. The optimized model achieved an overall MAE of 0.246 and an RMSE of 0.299.

Table 5.2: Performance of the Random Forest Regressor evaluated using Nested Leave-One-Out Cross-Validation.

Target	**MAE**	**RMSE**	R^2
Valence	0.207	0.270	0.186
Arousal	0.285	0.325	-0.023
Overall	0.246	0.299	0.082

As in the baseline evaluation, valence was predicted more accurately (MAE = 0.207, RMSE = 0.270, R^2 = 0.186) than arousal (MAE = 0.285, RMSE = 0.325,

R^2 = -0.023). The slightly lower performance compared to standard LOOCV is expected, as nested cross-validation provides a less biased estimate of generalization performance. Overall, the results indicate that the extracted drawing features are more informative for predicting valence than arousal.

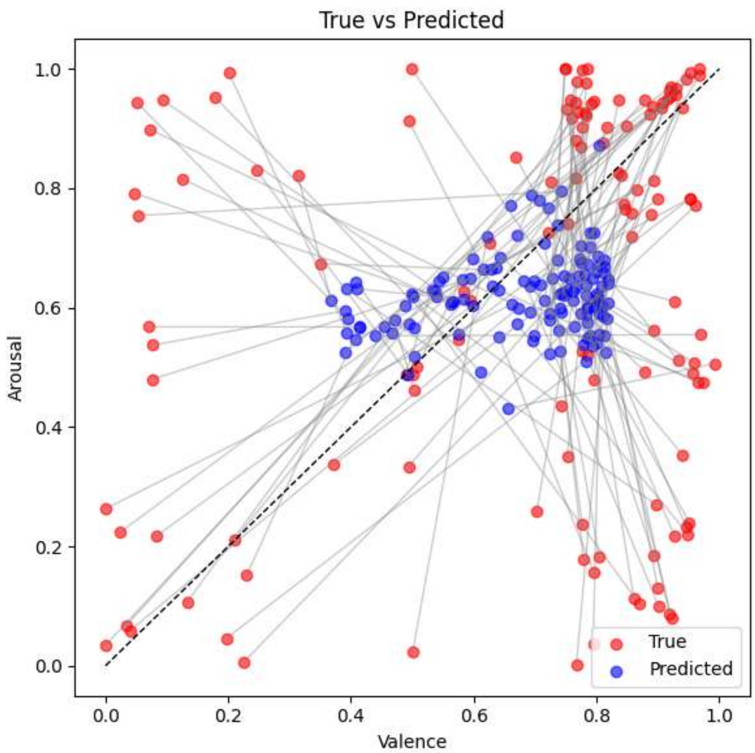


Figure 5.3: Cross-validation RMSE for different Random Forest hyperparameter configurations

Figure 5.3 compares the true and predicted valence-arousal annotations obtained with Nested LOOCV. The model captures the overall trend of the data but tends to predict values closer to the center of the emotional space, making extreme emotions harder to estimate. This behavior is likely influenced by the limited dataset and the uneven distribution of emotional annotations.

5.3 Randomized Search Results

The results obtained from the Randomized Search were used to define the hyperparameter search space for the subsequent Grid Search. Instead of exploring a large range of possible values, the Grid Search focused only on the most promising hyperparameter combinations identified during the Randomized Search, allowing a finer optimization of the Random Forest model.

Table 5.3: Hyperparameter search space used for the Grid Search, refined from the best hyperparameter values obtained during the Randomized Search.

Hyperparameter	**Values**
Number of trees (*n_estimators*)	30, 170, 190, 310
Maximum tree depth (*max_depth*)	None, 7
Minimum samples to split a node (*min_samples_split*)	2, 3
Minimum samples per leaf (*min_samples_leaf*)	1, 2
Maximum number of features (*max_features*)	1–10

5.3.1 Top N Feature Importance from the Randomized Search

Following the Randomized Search and the feature importance analysis, the ten most influential handcrafted features were selected. These features correspond to the highest average importance values obtained across all Leave-One-Out Cross-Validation (LOOCV) folds. They were subsequently used as the input features for the optimized Random Forest model and for the Random Forest component of the proposed hybrid model.

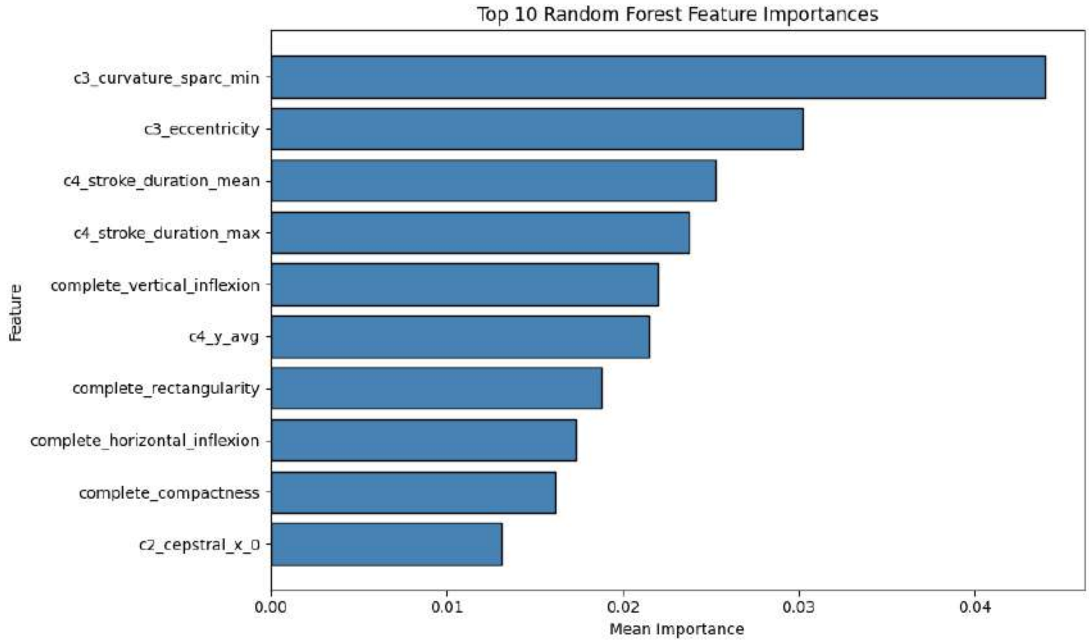


Figure 5.4: Top 10 Feature Importance from the Randomized Search

5.3.2 Train a Random Forest Grid Search on Top 10 Features and Best Parameters from the Randomized Search

After the Randomized Search identified the most promising hyperparameter values, a refined Grid Search was performed using only the 10 most important features. This reduced the feature space and enabled a more focused optimization of the Random Forest model.

Metric	**Overall**	**Valence**	**Arousal**
MAE	0.2283	0.1860	0.2706
RMSE	0.2850	0.2480	0.3177
R^2	0.1684	0.3135	0.0233

Table 5.4: Performance of the optimized Random Forest model using the top 10 most important features and Grid Search hyperparameter tuning.

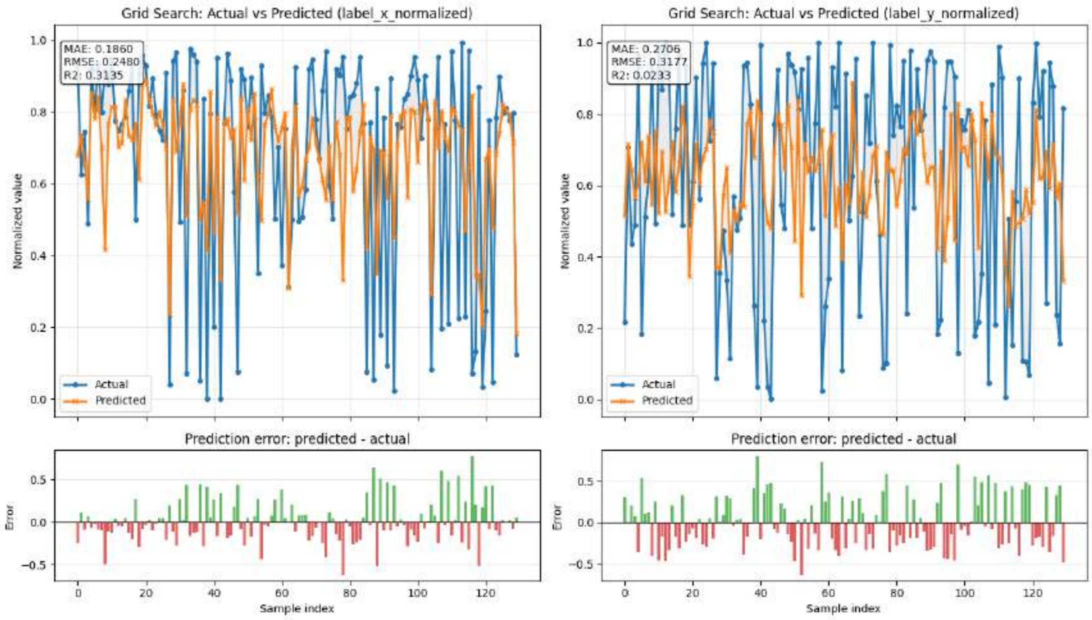


Figure 5.5: Prediction performance of the optimized Random Forest model using the top 10 selected features.

Figure 5.5 shows the performance of the optimized Random Forest model trained with the 10 most important features. Valence was predicted more accurately (MAE = 0.1860, RMSE = 0.2480, R^2 = 0.3135) than arousal (MAE = 0.2706, RMSE = 0.3177, R^2 = 0.0233). The predicted values follow the overall trend of the ground-truth annotations but tend to be less extreme. The error plots show that most prediction errors are centered around zero, although larger deviations are observed for arousal, highlighting its greater difficulty to predict.

5.4 Hybird Model Performance

Table 5.5 compares the performance of the Random Forest and MobileNetV2 models on the aligned dataset. The Random Forest consistently outperformed MobileNetV2, achieving lower MAE and RMSE values for both valence and arousal. It also achieved a positive R^2 for valence (0.3286), while MobileNetV2 produced negative R^2 values for both emotional dimensions. Overall, the hand-crafted trajectory features were more effective than the image-based approach, although the predictions from both models were retained for the subsequent hybrid fusion stage.

Metric	Model	Valence	Arousal
MAE	Random Forest	0.1890	0.2702
	MobileNetV2	0.2641	0.2872
RMSE	Random Forest	0.2518	0.3197
	MobileNetV2	0.3293	0.3428
R^2	Random Forest	0.3286	0.0078
	MobileNetV2	-0.1487	-0.1403

Table 5.5: Performance comparison between Random Forest and MobileNetV2 on the aligned hybrid dataset.

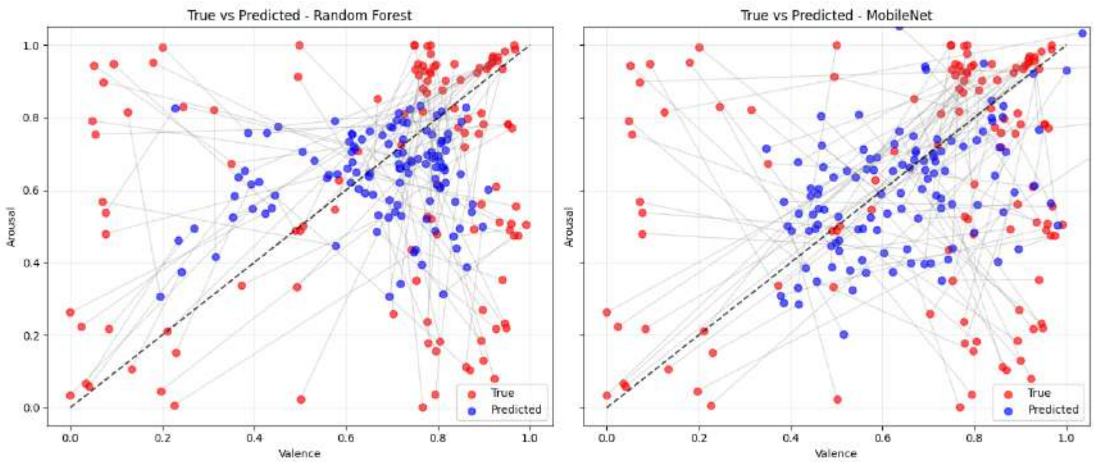


Figure 5.6: Comparison of the ground-truth and predicted emotional coordinates for the Random Forest and MobileNetV2 models

Figure 5.6 compares the predictions of the Random Forest and MobileNetV2 models in the valence-arousal space. The Random Forest predictions are generally closer to the ground-truth annotations, while MobileNetV2 shows larger deviations. These observations agree with the evaluation metrics, confirming that the Random Forest model provides more accurate emotion predictions than MobileNetV2 when the models are evaluated independently.

5.4.1 Hybird Model Results

The hybrid model was created by averaging the predictions of the Random Forest and MobileNetV2 models for both valence and arousal. The average fusion achieved an MAE of **0.2179** for valence and **0.2693** for arousal. The RMSE values were **0.2726** and **0.3154**, respectively. The model obtained an \$R2̂\$ of **0.2128** for valence and **0.0348** for arousal, showing that the hybrid approach improved arousal slightly, while valence remained better predicted by the Random Forest model.

Table 5.6: Performance of the average hybrid model on the aligned dataset.

Metric	**Valence**	**Arousal**
MAE	0.2179	0.2693
RMSE	0.2726	0.3154
R^2	0.2128	0.0348

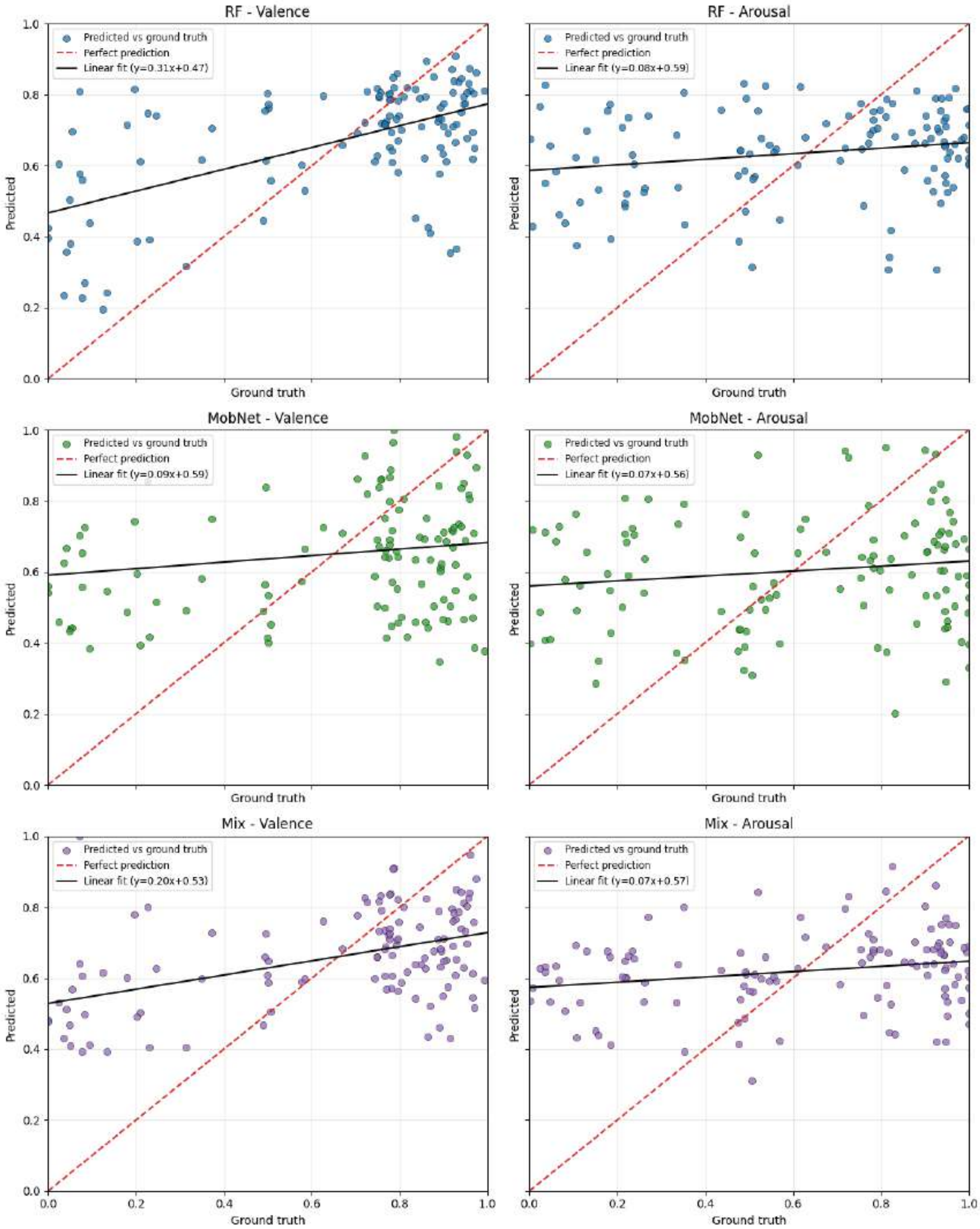


Figure 5.7: Scatter plots comparing the predicted and ground-truth valence and arousal values for the Random Forest, MobileNetV2, and hybrid models.

Figure 5.7 compares the predicted and ground-truth values for the Random

Forest, MobileNetV2, and hybrid models using scatter plots. The red dashed line represents perfect predictions, while the black line shows the linear trend between the predicted and actual values. A stronger relationship is indicated by points lying closer to the diagonal and a steeper regression line.

For valence, the Random Forest exhibits the strongest relationship with the ground truth, with predictions more closely following the ideal line. MobileNetV2 shows greater dispersion and a flatter trend, indicating weaker predictive performance. The hybrid model provides intermediate results but does not surpass the Random Forest.

For arousal, all three models display a much weaker relationship with the ground-truth values, as reflected by the nearly horizontal regression lines and the larger spread of points. This suggests that arousal is more difficult to predict than valence regardless of the model used.

Overall, the figure supports the quantitative evaluation, confirming that the Random Forest achieves the best overall performance, particularly for valence, while the hybrid model offers only a modest improvement over MobileNetV2 and limited gains compared with the Random Forest.

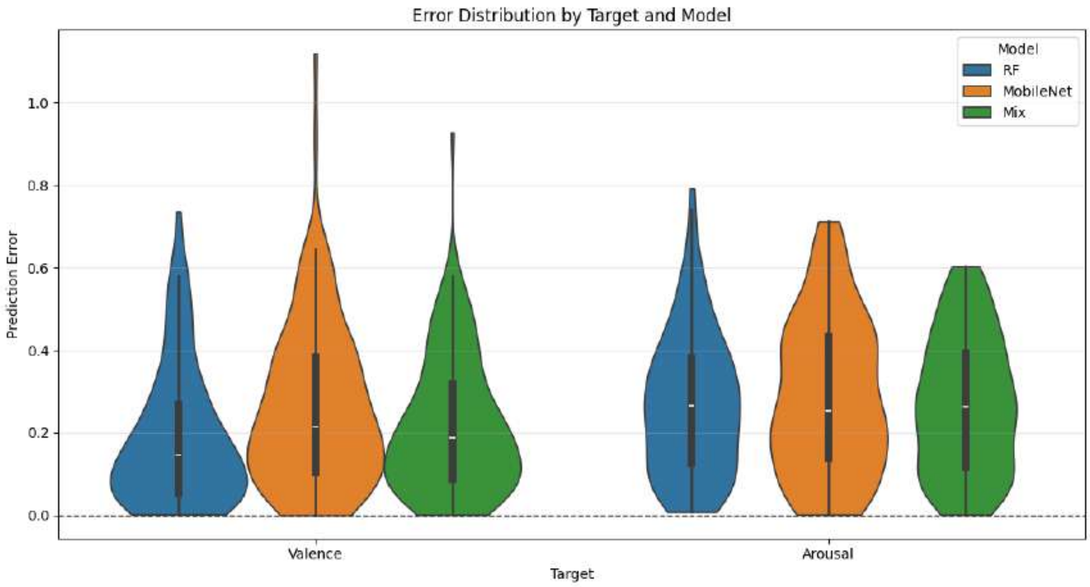


Figure 5.8: Error Distribution by Target and Model

Figure 5.8 presents the distribution of the absolute prediction errors for the Random Forest, MobileNetV2, and hybrid models on the valence and arousal dimensions. Since the errors are absolute, all values are non-negative, and smaller values indicate more accurate predictions. The width of each violin represents the density of the errors, while the embedded boxplot summarizes the median and interquartile range.

For valence, the Random Forest shows the lowest error distribution, with most prediction errors concentrated near zero. MobileNetV2 exhibits a wider distribution and larger errors, whereas the hybrid model produces intermediate results, reducing some of the larger errors but not outperforming the Random Forest. A similar pattern is observed for arousal, although the error distributions are generally wider for all three models, indicating that arousal is more difficult to predict than valence.

Overall, the figure supports the quantitative evaluation results. The Random Forest provides the most consistent predictions, particularly for valence, while the hybrid model offers only a modest improvement over MobileNetV2 and performs similarly to the Random Forest for arousal.

5.4.2 Wilcoxon Signed-Rank Test

The Wilcoxon Signed-Rank Test is a non-parametric statistical test used to compare two related sets of observations. It evaluates whether the median difference between paired samples is significantly different from zero. Unlike the paired t-test, the Wilcoxon test does not assume that the data are normally distributed, making it particularly suitable for small datasets or data that may violate the normality assumption.

In this study, each drawing produced a prediction from multiple models. Since the prediction errors from the Random Forest, MobileNetV2, and hybrid models were calculated on the same drawings, the observations are naturally paired. The Wilcoxon Signed-Rank Test was therefore used to compare the absolute prediction errors of the models while accounting for this pairing.

The test begins by computing the difference between the prediction errors of the two models for each drawing. Differences equal to zero are ignored because they provide no evidence that one model performs differently from the other. The remaining differences are converted to their absolute values and ranked from the smallest to the largest. The original sign of each difference (positive or negative) is then assigned to its corresponding rank. Finally, the ranks of the positive and negative differences are summed separately. If both models perform similarly, the positive and negative rank sums are expected to be nearly equal. A large imbalance between these sums indicates that one model consistently produces smaller prediction errors than the other.

The test returns two values: the Wilcoxon statistic (W) and the p-value. The Wilcoxon statistic summarizes the ranked differences between the paired observations, while the p-value indicates whether the observed difference is statistically significant. The hypotheses are defined as follows:

Null hypothesis (H_0): There is no significant difference between the prediction errors of the two models. Alternative hypothesis (H_1): There is a significant difference between the prediction errors of the two models.

A significance level of 0.05 was adopted in this study. If the p-value is less than 0.05, the null hypothesis is rejected, indicating that the difference in prediction errors between the two models is statistically significant. Otherwise, the null hypothesis cannot be rejected, suggesting that any observed difference is likely due to random variation.

The Wilcoxon Signed-Rank Test was performed separately for the valence and arousal dimensions to compare the performance of the Random Forest and hybrid models, as well as the MobileNetV2 and hybrid models. This analysis complements the evaluation metrics (MAE, RMSE, and R^2) by determining whether the observed differences in prediction accuracy are statistically meaningful rather than the result of chance.

Target	**Comparison**	W	p**-value**	$\alpha = 0.05$	**Decision**
Valence	RF vs Hybrid	2309	0.00083	Yes	Reject H_0
Arousal	RF vs Hybrid	3467	0.78475	No	Fail to reject H_0
Valence	MobileNet vs Hybrid	1620	2.33×10^{-7}	Yes	Reject H_0
Arousal	MobileNet vs Hybrid	2712	0.02289	Yes	Reject H_0

Table 5.7: Results of the Wilcoxon Signed-Rank Test comparing the prediction errors of the Random Forest, MobileNetV2, and hybrid models.

Table 5.7 presents the results of the Wilcoxon Signed-Rank Test used to compare the prediction errors of the hybrid model with those of the Random Forest and MobileNetV2 models. A significance level of **0.05** was adopted.

For the **Random Forest versus hybrid** comparison, a statistically significant difference was found for **valence** ($p = 0.00083$), indicating that the two models performed differently. Since the Random Forest achieved a lower MAE (0.1890) than the hybrid model (0.2179), the Random Forest provided significantly better predictions for valence. In contrast, no significant difference was observed for **arousal** ($p = 0.78475$), suggesting that the small difference in MAE between the two models (0.2702 vs. 0.2693) is not statistically meaningful.

For the **MobileNetV2 versus hybrid** comparison, statistically significant differences were found for both **valence** ($p = 2.33 \times 10$) and **arousal** ($p = 0.02289$). In both cases, the hybrid model achieved lower MAE values than MobileNetV2, indicating that combining the Random Forest and MobileNetV2 predictions significantly improved performance over using MobileNetV2 alone.

Overall, the Wilcoxon Signed-Rank Test confirms that the hybrid model provides a significant improvement over **MobileNetV2** for both emotional dimensions. However, it does **not** outperform the **Random Forest**. For valence, the Random Forest remains significantly more accurate than the hybrid model, while for arousal the two models achieve statistically similar performance. These findings are consistent with the MAE, RMSE, and $R2̂$ results presented earlier.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This thesis presented a multimodal framework for predicting children's emotions from drawings using machine learning techniques. The work extends the existing TRATTO tangible drawing interface by adding an automatic emotion recognition component capable of estimating children's affective states in the two-dimensional valence–arousal space.

Three different approaches were investigated. The first used a Random Forest Regressor trained on handcrafted features extracted from the drawing trajectories. These features described different aspects of the drawing process, including temporal, kinematic, geometric, structural, and spectral characteristics. The second approach employed a MobileNetV2 deep learning model trained directly on images of the completed drawings. Finally, a hybrid model combined the predictions of both approaches through a late-fusion strategy based on averaging.

The experimental evaluation showed that the Random Forest model achieved the best overall performance. After feature selection and hyperparameter optimization, the model produced the lowest prediction errors and the highest coefficient of determination, particularly for the valence dimension. The feature importance analysis also demonstrated that only a small subset of trajectory features contributed most of the predictive power, showing that the dynamics of the drawing process contain valuable information about children's emotional states.

Although the MobileNetV2 model was able to learn visual patterns from the completed drawings, its performance was lower than that of the Random Forest model. This suggests that, for the available dataset, the way children produced their drawings contained more useful emotional information than the final drawing itself.

The proposed hybrid model successfully combined the outputs of both models and achieved a small improvement over MobileNetV2, especially for arousal prediction. However, it did not outperform the optimized Random Forest model. The Wilcoxon Signed-Rank Test confirmed these findings by showing that the hybrid model significantly improved upon MobileNetV2 but did not provide a statistically significant improvement over the Random Forest model.

Overall, the results demonstrate that children's drawing behaviour contains meaningful information for estimating emotional states. In particular, trajectory-based features describing movement dynamics appear to be highly informative for continuous emotion prediction. These findings support the use of embodied interaction as a valuable source of affective information and demonstrate the potential of combining tangible interfaces with machine learning to support socio-emotional learning in educational environments.

Beyond the predictive models themselves, this work contributes a complete machine learning pipeline, including dataset preparation, feature extraction, model optimization, multimodal fusion, and statistical evaluation. The proposed framework provides a foundation for integrating automatic emotion recognition into tangible educational technologies, enabling teachers to better understand children's emotional expression while preserving the natural and embodied nature of the drawing activity.

6.2 Limitations

The main limitation is the relatively small size of the dataset. Although Leave-One-Out Cross-Validation and nested cross-validation were used to obtain reliable performance estimates, larger datasets would likely improve the generalization ability of the models, particularly for deep learning approaches.

Another limitation is the imbalance of the emotional annotations. Most drawings were concentrated in the high-valence, high-arousal region of the emotion space, while fewer samples represented other emotional states. This imbalance made some emotions more difficult to predict and likely contributed to the lower performance observed for arousal.

In addition, the hybrid model used a simple averaging strategy to combine the predictions of the Random Forest and MobileNetV2 models. More advanced fusion techniques could potentially exploit the complementary information from both modalities more effectively.

Finally, the experiments focused only on data collected with the TRATTO system in controlled educational activities. Additional validation in different classroom settings and with more diverse participant groups would further strengthen the applicability of the proposed framework.

6.3 Future Work

Several directions can be explored to further improve this research.

The first priority is the collection of a larger and more balanced dataset. Increasing the number of drawings and obtaining a more uniform distribution across the valence–arousal space would allow the models to learn a wider range of emotional behaviours and improve their ability to generalize.

Future work could also investigate more advanced multimodal fusion techniques. Instead of averaging the predictions of the individual models, feature-level fusion or attention-based deep learning architectures could better capture the complementary relationship between trajectory features and drawing images.

Another possible extension is the inclusion of additional modalities, such as pressure information, pen orientation, drawing timing, facial expressions, body movements, or physiological signals. Combining multiple behavioural sources may provide a richer representation of children's emotional states.

The framework could also be evaluated in long-term classroom studies to examine how automatic emotion recognition can support teachers during socio-emotional learning activities. Rather than providing isolated predictions, the system could generate emotional summaries over time, helping educators identify emotional trends, encourage discussion, and monitor classroom well-being.

Finally, future research should continue to address ethical aspects of affective computing in educational environments. Ensuring transparency, protecting children's privacy, and supporting teachers rather than replacing their judgment remain essential considerations for the responsible deployment of intelligent educational technologies.

References

[1] J. A. Durlak, R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger, “The impact of enhancing students’ social and emotional learning: A meta-analysis of school-based universal interventions,” *Child Development*, vol. 82, no. 1, pp. 405–432, 2011.

[2] C. S. Bailey, O. Martinez, and E. DiDomizio, “Social and emotional learning and early literacy skills: A quasi-experimental study of ruler,” *Education Sciences*, 2019. Yale Child Study Center, Yale School of Medicine, Yale University.

[3] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, “Multimodal analysis of expressive gesture in music and dance performances,” in *Gesture-Based Communication in Human-Computer Interaction* (A. Camurri and G. Volpe, eds.), vol. 2915 of *Lecture Notes in Computer Science*, (Berlin, Heidelberg), Springer, 2004.

[4] L. Shapiro and S. A. Stolz, “Embodied cognition and its significance for education,” *Theory and Research in Education*, vol. 17, no. 1, pp. 19–39, 2019.

[5] S. Ferrando, N. Corbellini, G. Lepri, G. Volpe, and E. Ceccaldi, “Colour me happy: towards a tangible interface system for embodied socio-emotional

learning in children," in *Proceedings of the University of Genoa DIBRIS Research Workshop*, (Genova, Italy), Casa Paganini, InfoMus, DIBRIS.

[6] Y. Lin, G. Yang, Y. Ze, L. Zhang, B. Xing, X. Liu, and R. Lyu, "The impact of motion features of hand-drawn lines on emotional expression: an experimental study," *Computers & Graphics*, vol. 119, p. 103897, 2024.

[7] N. Eisenberg, W. Damon, and R. M. Lerner, *Handbook of Child Psychology, Social, Emotional, and Personality Development*, vol. 3. Hoboken, NJ: John Wiley & Sons, 6th ed., 2006.

[8] K. Weare and M. Nind, "Mental health promotion and problem prevention in schools: what does the evidence say?," *Health Promotion International*, vol. 26, no. S1, 2011.

[9] R. Bar-On, K. Maree, and M. J. Elias, eds., *Educating People to Be Emotionally Intelligent*. Westport, CT: Praeger Publishers, 2007. Foreword by Daniel Goleman.

[10] M. T. Greenberg, "School-based prevention: Current status and future challenges," *Effective Education*, vol. 2, no. 1, pp. 27–52, 2010.

[11] J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, and T. P. Gullotta, eds., *Handbook of Social and Emotional Learning: Research and Practice*. New York, NY: Guilford Press, 2015.

[12] J. A. Durlak, R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger, "The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions," *Child Development*, vol. 82, no. 1, pp. 405–432, 2011.

[13] M. A. Brackett and S. E. Rivers, "Transforming students' lives with social and emotional learning," in *Handbook of Emotions in Education* (R. Pekrun and L. Linnenbrink-Garcia, eds.), pp. 368–388, New York, NY: Routledge, 2014.

[14] M. A. Brackett, M. R. Reyes, S. E. Rivers, N. A. Elbertson, and P. Salovey, "Developing emotion skills in early childhood settings using preschool ruler," *Psychology in the Schools*, vol. 52, no. 1, pp. 62–77, 2015.

[15] C. Hagelskamp, M. A. Brackett, S. E. Rivers, and P. Salovey, "Improving classroom quality with the ruler approach to social and emotional learning: Proximal and distal outcomes," *American Journal of Community Psychology*, vol. 51, no. 3-4, pp. 530–543, 2013.

[16] C. C. Raver, P. W. Garner, and R. Smith-Donald, "The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal?," in *School readiness and the transition to kindergarten in the era of accountability* (R. C. Pianta, M. J. Cox, and K. L. Snow, eds.), pp. 121–147, Baltimore, MD: Paul H. Brookes Publishing Co., 2007.

[17] N. D. Eggum, N. Eisenberg, K. Kao, T. L. Spinrad, R. Bolnick, C. Hofer, A. S. Kupfer, and W. V. Fabricius, "Emotion understanding, theory of mind, and prosocial orientation: Relations over time in early childhood," *The Journal of Positive Psychology*, vol. 6, no. 1, pp. 4–16, 2011.

[18] J. E. Zins, R. P. Weissberg, M. C. Wang, and H. J. Walberg, eds., *Building Academic Success on Social and Emotional Learning: What Does the Research Say?* New York, NY: Teachers College Press, 2004.

[19] L. W. Barsalou, "Grounded cognition," *Annual Review of Psychology*, vol. 59, pp. 617–645, 2008.

[20] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind, revised edition: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, revised ed., 2017.

[21] L. Shapiro and S. A. Stolz, "Embodied cognition and its significance for education," *Theory and Research in Education*, vol. 17, no. 1, pp. 19–39, 2019.

[22] M. Macedonia, A. Groher, and F. Roithmayr, "Implementing embodied learning in the classroom: effects on children's memory and hands-on skills," *Educational Psychology Review*, vol. 31, no. 2, pp. 507–530, 2019.

[23] N. H. Frijda, "Emotion, cognitive structure, and action tendency," *Cognition & Emotion*, vol. 1, no. 2, pp. 115–143, 1987.

[24] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Series in Affective Science, New York, NY: Oxford University Press, 2004.

[25] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 213–225, 2003.

[26] A. Camurri, G. Volpe, S. Piana, M. Mancini, R. Niewiadomski, N. Ferrari, and C. Canepa, "The dancer in the eye: Towards a multi-layered computational framework of qualities in movement," in *Proceedings of the 3rd International Symposium on Movement and Computing (MOCO '16)*, (New York, NY, USA), Association for Computing Machinery (ACM), July 2016.

[27] O. Shaer and E. Hornecker, "Tangible user interfaces: Past, present, and

future directions," *Foundations and Trends® in Human–Computer Interaction*, vol. 3, no. 1-2, pp. 1–137, 2010.

[28] E. Hornecker and J. Buur, "Getting a grip on tangible interaction: a framework on physical space and social interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, (New York, NY, USA), pp. 437–446, Association for Computing Machinery (ACM), 2006.

[29] D. Abrahamson, "Embodied design: constructing means for constructing meaning," *Educational Studies in Mathematics*, vol. 70, no. 1, pp. 27–47, 2009.

[30] P. Slovák and G. Fitzpatrick, "Teaching and developing social and emotional skills with technology," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 22, no. 4, pp. 1–34, 2015.

[31] M. Rizvi, "Designing tangibles for children's sel in conversations: Why and how," in *IS-EUD 2017: 6th International Symposium on End-User Development (Extended Abstracts)*, (Eindhoven, Netherlands), pp. 100–103, Eindhoven University of Technology, 2017.

[32] L. M. Triona, D. Klahr, and C. Williams, "Point and click or build by hand: Comparing the effects of physical vs. virtual materials on middle school students' ability to optimize an engineering design," in *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, (Mahwah, NJ), pp. 1170–1175, Lawrence Erlbaum Associates, 2003.

[33] P. Marshall, "Do tangible interfaces enhance learning?," in *Proceedings of the 1st International Conference on Tangible and Embedded Interaction (TEI*

'07), (New York, NY, USA), pp. 163–170, Association for Computing Machinery (ACM), 2007.

[34] A. Skulmowski and G. D. Rey, "Embodied learning: introducing a taxonomy based on bodily engagement and task integration," *Cognitive Research: Principles and Implications*, vol. 3, no. 1, p. 6, 2018.

[35] H. Lundholm, "The affective tone of lines: Experimental researches," *Psychological Review*, vol. 28, no. 1, pp. 43–60, 1921.

[36] A. T. Poffenberger and B. E. Barrows, "The feeling value of lines," *Journal of Applied Psychology*, vol. 8, no. 2, pp. 187–205, 1924.

[37] Y. Hu, R. Lyu, and X. Liu, "Is the emotional mapping of lines caused by the motion they imply?," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43, 2021.