Signature-based manual dating vs. neural network automation

Tea Tvalavadze[1,†], Ia Ghadua[1,†], Giorgi Kalandadze[2,†] and Maksim Iavich[3,*,†]

[1] *Giorgi Leonidze State Museum of Georgian Literature, 36 Petre Kavtaradze str., Tbilisi, 0186, Georgia*

[2] *Association for Textual and Editorial Studies and Digital Humanities, 17 Sakhalkho str., Tbilisi, 0113, Georgia*

[3] *Caucasus University, School of Technology, 1 Paata Saakadze str., Tbilisi, 0102, Georgia*

Abstract

Dating manuscripts is a multifaceted task that necessitates integrating various analytical methods to establish historical context and authenticate documents. This paper compares two methods for dating manuscripts of Galaktion Tabidze, a notable Georgian poet. We utilized both a manual signature analysis and an automated Convolutional Neural Network (CNN) approach to date undated manuscripts from Tabidze's archive. The manual signature method relied on analyzing specific graphematic features associated with different periods of Tabidze's work. This approach provided clear and consistent dating results for manuscripts. The CNN method, on the other hand, used probabilistic estimates to suggest dates. While the CNN method generally supported the manual findings, it also introduced some uncertainties. For instance, the CNN method suggested certain dates that did not align with the manual analysis, such as late 20th-century dates for manuscripts that the manual method dated to earlier periods. The comparison highlighted that the manual signature method offered more reliable and precise dating, especially for earlier works. The CNN method, while valuable, introduced variability and indicated areas where the model's accuracy could be improved. This study demonstrates that while both methods have their strengths, the manual approach provides a more consistent basis for dating manuscripts, whereas the CNN method serves as a complementary tool with potential for further refinement.

Keywords

manuscript dating, neural networks, manual dating, automatic dating

1. Introduction

Dating manuscripts is a complex and multifaceted task that requires careful analysis and the integration of various methods. This process is essential for understanding a manuscript's historical context and verifying its authenticity. Several approaches are used: Examining physical characteristics such as paper, ink, and binding materials can provide clues about the manuscript's age. Specific types of paper and ink, along with features like watermarks and script styles, can often be linked to particular periods or regions [1, 2]. Paleography, or the study of ancient handwriting styles, is another critical method. By analyzing the script, scholars can identify changes in handwriting over time, which helps in establishing the manuscript's timeframe. Historical references within the manuscript—such as mentions of events, figures, or other works—can offer dating clues. If the manuscript refers to specific historical events or individuals with known dates, this information can help narrow down its creation period. Textual analysis, which involves comparing the manuscript's content with other known works, also plays a role [3–6]. Quotations or influences from texts with established dates can assist in dating the manuscript. For manuscripts on organic materials like paper or parchment, carbon dating can estimate the age of the material. Although this provides a date range, it may not pinpoint the exact date of the manuscript's creation. The manuscript's provenance, or its history of ownership, can offer additional dating information. Inscriptions, ownership marks, or historical records related to previous owners can provide valuable clues [7, 8].

In the initial phase of the graphematic analysis of the manuscripts of Galaktion Tabidze, the preeminent Georgian poet of the 20th century (1891–1959), the research team aimed to date the manuscripts preserved in his archive. The team selected 2–3 dated manuscripts from each year between 1905 and 1959. They deconstructed the scanned images, compiled databases of graphemes and graphemic pairs, and identified the most informative element types for dating purposes, subsequently coding these elements [9, 10]. The database of undated manuscripts was then processed using the same principles,

0000-0003-3742-6825 (T. Tvalavadze); 0000-0002-7434-2519 (I. Ghadua); 0009-0002-3269-0133 (G. Kalandadze); 0000-0002-3109-7971 (M. Iavich)

with an attempt to date them based on specific graphemic features identified over the years. The dating of the test manuscripts revealed that the predominant presence of the 622 graphemic types across all periods of Tabidze's work hindered precise dating, thus impacting the overall results [11].

Given that a single comparative analysis of all the features of all graphemes did not yield significant results, we decided to refine our approach by focusing on specific types of elements that were either consistently used or distinctly absent during particular periods. To achieve this, we conducted a detailed examination of each of the 622 graphematic types identified in our codebook. Our findings indicated that most types appeared in manuscripts from specific years, but not in adjacent years, and then reappeared after a gap of one or two years. This pattern likely resulted from random variation rather than date-informative differences between manuscripts from these years. Nonetheless, the extended periods of use or non-use of specific types revealed through this study could provide a robust basis for more accurate dating. In this paper we present two methods of dating the manuscripts, the first one, the manual using the signatures of the authors, and the second one using the Neural Networks.

2. Dating the manuscripts using signature methods

As previously mentioned, the majority of the graphemic types we identified were present in manuscripts from at least one or two years across each decade of the poet's activity: the 1910s, 1920s, 1930s, 1940s, and 1950s. Consequently, relying solely on these bases made it difficult to accurately date the texts. However, we encountered several exceptions. For instance, the analysis of dated manuscripts revealed that the element type >ჳ<2/2] was used by the poet exclusively from April 1908 to December 1910. This discovery allowed us to date dozens of undated manuscripts that utilized this specific type.

During the compilation of the codebook, we noted the presence of various graphemic types in the manuscripts of each year but did not record the percentage relationship of each type with other types of the same element. Consequently, negative statistics—identifying which types were absent in specific years—proved more fruitful than positive statistics in our research. In other words, understanding which types were not found in certain years provided more valuable insights. Given that the earliest extant manuscript of Galaktion dates from 1905 and the latest from 1959, element types absent from 1905–1915 suggest these elements began to be used in 1916, allowing us to date manuscripts containing these types to periods post-1915. Similarly, element types present between 1949 and 1959, but absent in earlier periods, indicate that the poet ceased using these types after 1949, enabling us to date texts with these types to before 1949. Naturally, a conclusive determination cannot be based on the presence or absence of a single grapheme. Therefore, we also conducted studies on other graphemes to confirm their compatibility with the estimated periods we identified.
As previously mentioned, extended time intervals indicating the use or non-use of specific types of elements cannot independently solve the problem of dating a text.

However, they can contribute significantly to achieving accurate results within a comprehensive research framework. For instance, in cases where multiple potential dates have been identified through other methods, these intervals can help us choose the most likely date.

For example, let's consider the case of one poem by Galaktion Tabidze, which he published in 1933 under the title "The First of May" with the inscription "Poem delivered at an illegal evening in 1908, on the first of May". The date 1908 is written not only on the publication but also on the autographs. The fact is that Galaktion regularly published his poems, and, unless some external circumstances prevented it, nothing remained unpublished. Given the social democrats' rise to power in Georgia in 1918 and the subsequent annexation by Soviet Russia in 1921, it is clear that there would have been no obstacles to publishing this poem from 1918 onwards. Consequently, there is suspicion that the poet may have attributed a false date to the poem to construct an image of his "revolutionary past". Considering the circumstances of writers under Stalin's totalitarian regime, this is not surprising. While no opposition has arisen since the collapse of the Soviet Union to the notion that Galaktion might have falsified the date, definitive proof confirming the falsification of this poem's date by the author has not yet emerged.

Five of the six autographs of the poem are dated: two indicate 1905, two indicate 1908, and one, at the end of the poem, is dated April 26, 1933. Such variability in the dating by the author naturally strengthened the suspicion of its mystification. However, graphematic research provided an opportunity to substantiate this assumption. Since three autographs of the poem (MGL, N4763, N5381, N5552) are corrected so much that they reflect the process of creating the poem rather than merely "copying" it, it was evident that determining the creation time of these manuscripts through graphemic analysis would aid in pinpointing the poem's actual date of composition. In this instance, our task was relatively straightforward: we needed to select one of the three possible dates inscribed on the manuscripts. Given that the handwriting from 1905–1908 shares all the characteristics required for our common research parameters, our choice was effective between these dates and 1933. Here, there were numerous distinguishing features to consider.

Until 1911, no type of the double-arched "ღ" (>ღ<3/) is found in the Galaktion's manuscripts. The types of the double-arched "ჩ" (>ჩ<4/) and the upper additional line of "ლ" (>ლ<6/) do not appear until 1909, the type of grapheme "მ" (>მ<2/7])—before 1910, type of "წ" (>წ<1/9]) until 1912, type of "ვ" (>ვ<3/4])—until 1915, etc. Therefore, the presence of all these elements in the manuscripts of the poem "The First of May", and in large numbers, is a factual confirmation that they were created in 1933, and not in 1905 or 1908. See the Fig. 1.
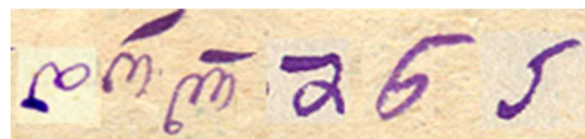

Figure 1: Elements set

The primary basis for dating manuscripts through graphematic research lies in the systematic variation of outlines over the years. People, especially writers, often

modify the outlines of individual graphemes and their tied pairs. However, as our experience indicates, they tend to focus more extensively on their facsimiles, which are directly linked to their individuality. Consequently, at the next stage of our graphematic research, we decided to observe and analyze the facsimiles of Galaktion Tabidze.

From one of the poet's recollections, we learn about his keen observation of his singing teacher's signing process at the Kutaisi theological school: "Sharabi-dze: for this "dze" he would draw a fast first line, then he would turn it with a second line, then he would add a third line. These were musical score lines, and on these lines, he would draw the clef so quickly and beautifully that I was amazed..." It appears that he admired the teacher's signature style so much that he developed a similar signature himself. Initially, he depicted the violin key horizontally in the lower part of the signature (D-273) and later began to shape the initials of his name and surname into a vertical violin key. He greatly appreciated when the initials of the name and surname, or ideally, their syllables, were repeated. See Fig. 2.

Figure 2: Signature basic

Several pages are filled with the Russian signatures of Ieronim Yasinsky, in which he first altered the initial of the name to a letter similar to the initial of the surname, then changed the surname ending to a feminine one, included another syllable containing the initial in the surname itself, and invented a name that would begin with the same syllable as the surname. Observing these practices aided us in analyzing the variations he introduced into his facsimile. See Fig. 3.
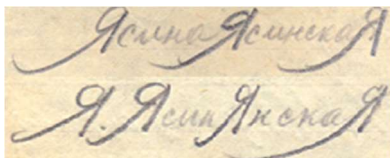

Figure 3: Russian signature

Galaktion Tabidze was highly sensitive to the issue of authorship, experiencing great distress even when others appropriated individual rhymes and not the entire poems. Consequently, it became his custom to sign each poem at the end, even when he wrote dozens of poems in a single notebook. This practice provided us with a wealth of material for graphematic research. We began by collecting all facsimiles from the poet's extensive archival units, sorting the facsimiles of dated manuscripts by year, and creating a separate database for updated ones. Our research commenced with the study of dated facsimiles, identifying the constituent elements of each facsimile and categorizing the types within these elements. In this analysis, the types

of outlines that characterized the author's facsimiles during specific periods were particularly valuable for dating. Conversely, those that appeared almost every year or only once or twice were deemed ineffective for this purpose.

Galaktion began publishing poems at the age of 17, in 1908, and his exceptional creative potential became immediately apparent. As he was an aesthete by nature, it was very important for him what kind of facsimile would appear under the autographs of his poems. Therefore, he invested substantial effort into perfecting it. His notebooks reveal numerous facsimiles written consecutively and reflect the meticulous process of working on and refining his facsimile. In one case, his signature took the shape of a ship and in another, it had the contours of an ornament. The graphemes were alternately enlarged, elongated, or angular, and sometimes the initials of his name and surname were either combined or intricately inserted one into the other. See Fig. 4.

**Figure 4:** Signature with initials

The signatures preserved in the poet's archive exhibit various compositional forms: the full first and last name, the first name abbreviated ("Gal.") with the full last name, the first name initial with the last name abbreviated in the middle ("G. T-dze"), first and last name initials ("G. T."), the first name only ("Galaktion"), the abbreviated first name only ("G."), and the first name initial with the final letters of the last name unidentifiable due to the rapid writing style. Among these, the latter type, the typical automated facsimile, demonstrates the most variations over the years. This type aims to indicate authorship rather than to perfectly represent all graphemes.

Galaktion appears to have been particularly fond of incorporating symmetrical elements into his facsimiles. As previously discussed, he created a symmetrical signature for the surname Yasinsky. Regarding his facsimiles, starting in 1907, he began adding bold horizontal lines to the initials of his first and last names to emphasize symmetry. During 1908–1909, he sometimes combined these two horizontal lines into one. Please see Fig. 5.

Figure 5: Signature with symmetry

By 1910, he started to write in bold the upper and lower parts of the vertical "curly" element, appended to the right side of the facsimile so that they were symmetric to the horizontal lines, added to the letter "g" or "g" and "t". Such facsimiles can be found up to and including 1915. Please see Fig. 6.
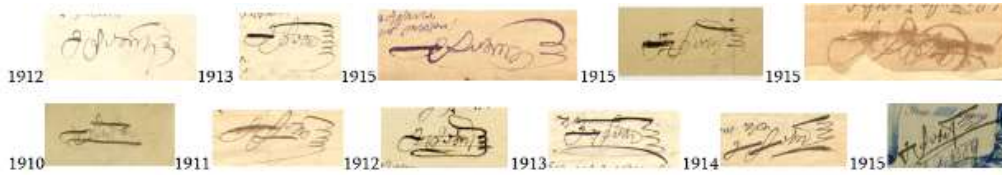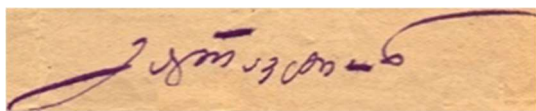
Figure 6: Signature with bold characters

Galaktion's desire to incorporate symmetry into his facsimiles persisted even beyond 1915. However, starting in 1916, he abandoned the use of vertical "curl" and instead introduced symmetry by accentuating the individual arc forms of the graphemes in the facsimile. For instance, the arc opening to the right of the "t" and the arc opening to the left of the "e" framed the facsimile. When the facsimile featured the initial of his name, this mirror symmetry was created by the arc of the "g" rather than the "t". See Fig. 7.

Figure 7: Signature with accentuating characters

From 1917 onwards, the poet began writing his full name, "Galaktion", and emphasized symmetry by darkening the additional line of the "l", the horizontal line of the "o," and sometimes the upper part of the "n". See Fig. 8.

Figure 8: Signature with full name.

Accordingly, the comparative study of the chronological database of Galaktion's facsimiles revealed four distinct time intervals characterized by specific stylistic elements. From 1907 to 1910, he utilized the symmetry of the letters "g" and "t" by adding horizontal lines to them. Between 1910 and 1915, he enhanced this symmetry by darkening the top, bottom, or both top and bottom lines of a vertical "curl". Starting in 1916, he emphasized the left part of his first name initial and the outlines "t", and "e" in his last name. Since 1917, he introduced symmetry through the upper part of the italicized "g", the additional line of the "l", the horizontal line of the angular "o", and the upper arc of the "n".

Facsimiles similar to the aforementioned first feature, identified through the research of the dated text facsimiles, were not found in the updated database. However, based on the discovery of the second feature, we dated a number of the archival units to the years 1910–1915, specifically: MGL: 45, 416, 650, 651, 1330, 1319–1327, 1361, and 2176. The third feature allowed us to date archival units to the period after 1916, including MGL 471−19, 507−2, 527−6, 603−12, 618−1, 637−8, 655−2, 1378, and 24551−242. The fourth feature indicated a period after 1917 for the following archival units: MGL 420−8, 488−1, 496−3, 638−10, 1378, 1678, 1718, and

3819. As a result, we were able to date 22 verses, 1 poem, 1 play, 14 diaries, and 2 personal letters belonging to Galaktion Tabidze. Although in some cases we could only determine the lower limit of the time interval, this is still significant. Without at least an approximate date, it would be impossible to include these works in the bio-bibliography and place them accurately within the author's biographical context.

3. Neural networks for manuscript dating

Neural networks have emerged as powerful tools in the analysis of historical manuscripts, offering innovative methods for extracting and interpreting various attributes of these documents. By leveraging advanced machine learning techniques, researchers can gain new insights into the content, structure, and material characteristics of manuscripts. The process begins with the collection of high-resolution images of manuscript pages, capturing detailed information about the text, handwriting, paper quality, and ink composition. Annotated datasets, comprising manuscripts with known attributes, serve as a foundation for training neural networks. These datasets provide the necessary ground truth for the network to learn from and refine its analysis capabilities [12–15].

Neural networks, particularly Convolutional Neural Networks (CNNs), excel in feature extraction from image data. For text analysis, Optical Character Recognition (OCR) is employed to convert handwritten text into machine-readable formats, enabling the network to identify and analyze text patterns and stylistic elements. In addition to text, neural networks can examine handwriting styles, detecting variations and trends that may indicate authorship or stylistic changes over time [16–19]. They also assess the physical attributes of manuscripts, such as paper texture, ink fading, and unique markings, which provide insights into the materials and methods used during the manuscript's production. Training the neural network involves using these annotated datasets to help the model recognize and learn patterns and features associated with different manuscript attributes. Advanced models, including Recurrent Neural Networks (RNNs) or Transformers, can be used for sequential text analysis, while transfer learning allows for the use of pre-trained models to enhance accuracy.

Once trained, the neural network can be applied to new manuscripts to perform detailed analyses. It can identify patterns, detect subtle variations, and extract meaningful features from both the text and physical attributes of the manuscript. This capability extends to recognizing specific types of paper or ink, assessing text style, and detecting unique annotations or markings. Validation of the neural

network's predictions involves cross-referencing with known samples and expert evaluations to ensure accuracy and reliability. Integrating the outputs of neural networks with expert knowledge is crucial, as it provides a comprehensive understanding of the findings and their historical context.

The advantages of using neural networks in manuscript analysis include their ability to provide objective, data-driven insights and efficiently process large volumes of data. They are capable of identifying intricate patterns and features that might elude traditional methods and can be scaled to analyze manuscripts from diverse historical periods and regions. However, there are challenges to consider. The quality of the training data is critical, and obtaining high-quality, annotated datasets can be demanding. Manuscripts often exhibit complex and overlapping features, which require advanced neural network architectures to model effectively. Additionally, neural network findings should be validated with expert input to ensure that interpretations align with historical knowledge. Looking forward, integrating neural networks with other analytical techniques, such as chemical analysis or historical records, can further enhance manuscript analysis. Expanding datasets to cover a broader range of historical contexts will improve model generalization, and ongoing advancements in neural network technology promise to refine and enhance analytical capabilities.

Neural networks represent a significant advancement in manuscript analysis, offering new methodologies for understanding and authenticating historical documents. Their application enables a more detailed examination of text and physical attributes, contributing to a deeper comprehension of manuscript origins and characteristics. The next section describes the methodology of using neural networks in manuscript dating.

4. The automatic method using neural networks

The manuscript dataset resides in the good drive folder `/content/drive/MyDrive/photos`, structured into subdirectories corresponding to different historical periods. This organizational scheme allows TensorFlow's `ImageFolder` utility to efficiently load and categorize images based on their respective periods. By leveraging directory names as class labels, the dataset loading process is streamlined, facilitating subsequent preprocessing steps.

Upon loading, the dataset is split into training and validation sets using `image_dataset_from_directory`. This function partitions the dataset based on a specified validation split (in this case, 5%), ensuring that a small portion of data is reserved for model validation. Parameters such as image size and batch size are configured to standardize input dimensions (`img_height` and `img_width` set to 180 pixels each) and optimize memory usage during training.

Data augmentation is implemented using TensorFlow's `RandomFlip` method, which introduces variations in training images by randomly flipping them horizontally and vertically. This technique is crucial for enhancing model robustness and generalization, as it exposes the model to diverse perspectives and orientations of manuscript images.

The CNN model architecture is defined sequentially using Keras, starting with a `Rescaling` layer to normalize pixel values between 0 and 1. Normalization ensures consistency in input data, facilitating efficient model convergence during training. The subsequent `Sequential` container encapsulates layers responsible for feature extraction and classification.

The model incorporates three convolutional layers (`Conv2D`), each followed by a `MaxPooling2D` layer. Convolutional layers apply a set of filters to extract hierarchical features from input images, while max-pooling layers downsample feature maps, reducing computational complexity and focusing on prominent features. These operations enable the model to learn spatial hierarchies and abstract representations inherent in manuscript images.

Batch normalization layers (`BatchNormalization`) are interspersed between convolutional and activation layers, stabilizing training by normalizing activations and accelerating convergence. This technique enhances model training efficiency and robustness to variations in input data.

Following convolutional operations, feature maps are flattened (`Flatten`), converting multi-dimensional tensors into one-dimensional vectors suitable for dense layers. Two fully connected (`Dense`) layers with ReLU activation functions facilitate nonlinear mapping and feature aggregation. The first dense layer employs L2 regularization (`regularizers.l2`) to mitigate overfitting, penalizing large weights, and promoting model generalization.

The output layer consists of `num_classes` units corresponding to the number of historical periods in the dataset. Utilizing softmax activation, the output layer computes probabilities for each period, facilitating multi-class classification by assigning manuscripts to their most likely historical categories based on learned features.

To optimize model parameters, the model is compiled with the Adam optimizer, known for its efficiency in stochastic optimization tasks. Sparse categorical cross-entropy serves as the loss function, appropriate for multi-class classification where each manuscript is assigned a single historical period label. Training commences using the `fit` method, iterating over a specified number of epochs (30 epochs in this case) to adjust model weights based on training data (`train_ds`). Validation against separate validation data (`val_ds`) assesses model performance on unseen examples, preventing overfitting and validating its ability to generalize to new manuscripts. During training, metrics such as accuracy and loss are monitored and visualized using matplotlib. Plots of training/validation accuracy and loss across epochs provide insights into model convergence and performance trends, aiding in the assessment of model efficacy and identification of potential improvements.

Upon completing training, the trained model is saved using `model.save`, storing the model's architecture, weights, and optimizer state on disk. This step ensures that the trained model can be reused and deployed for inference on new manuscript images without the need for retraining.

In addition to model saving, a zip archive containing model files (`mymodel.zip`) is created, enhancing

portability and facilitating distribution for collaborative research or deployment in digital archives.

For inference on new, undated manuscripts, the saved model is loaded using `tf.keras.models.load_model`. Manuscript images from a designated directory (`drive/MyDrive/toCheck`) are loaded and preprocessed using TensorFlow's image processing utilities. Each image undergoes resizing (`target_size = (180, 180)`) and conversion into a numerical format suitable for input to the model.

Inference is conducted using `model.predict`, generating predictions in the form of probabilities for each historical period. The top three predicted periods for each manuscript are stored in a dictionary (`resultDict`), facilitating further analysis and validation by historians and researchers. This approach enables automated dating and categorization of undated manuscripts based on visual content, leveraging machine learning to support historical research and analysis.

The pseudo code of CNN:

```
# Step 1: Data Handling and Preprocessing

# Define data directory
data_dir = '/content/drive/MyDrive/photos'

# Load dataset using ImageFolder
builder = tfds.ImageFolder(data_dir)
dataset = builder.as_dataset(shuffle_files=True)

# Split dataset into training and validation sets
train_ds = tf.keras.utils.image_dataset_from_directory(
  data_dir,
  validation_split=0.05,
  subset="training",
  seed=123,
  image_size=(180, 180),
  batch_size=32
)

val_ds = tf.keras.utils.image_dataset_from_directory(
  data_dir,
  validation_split=0.05,
  subset="validation",
  seed=123,
  image_size=(180, 180),
  batch_size=32
)

# Apply data augmentation
data_augmentation = keras.Sequential([

layers.experimental.preprocessing.RandomFlip("horizontal_and_vertical"),
  ])

# Normalize pixel values
normalization_layer = tf.keras.layers.Rescaling(1./255)

# Prepare dataset for training and validation
train_ds = train_ds.map(lambda x, y: (normalization_layer(x), y))
val_ds = val_ds.map(lambda x, y: (normalization_layer(x), y))

# Cache and prefetch datasets for performance
AUTOTUNE = tf.data.AUTOTUNE
train_ds = train_ds.cache().prefetch(buffer_size=AUTOTUNE)
val_ds = val_ds.cache().prefetch(buffer_size=AUTOTUNE)


# Step 2: Model Architecture

# Define CNN model architecture
model = tf.keras.Sequential([
  tf.keras.layers.Rescaling(1./255),
  data_augmentation,
  tf.keras.layers.Conv2D(32, (3, 3), activation='relu'),
  tf.keras.layers.MaxPooling2D((2, 2)),
  tf.keras.layers.Conv2D(64, (3, 3), activation='relu'),
  tf.keras.layers.MaxPooling2D((2, 2)),
  tf.keras.layers.Conv2D(128, (3, 3), activation='relu'),
  tf.keras.layers.MaxPooling2D((2, 2)),
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(128, activation='relu', kernel_regularizer=regularizers.l2(0.001)),
  tf.keras.layers.Dense(num_classes)  # num_classes is the number of historical periods
])

# Compile the model
model.compile(
  optimizer='adam',
  loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
  metrics=['accuracy']
)


# Step 3: Model Training and Evaluation

# Train the model
history = model.fit(
  train_ds,
  validation_data=val_ds,
  epochs=30
)

# Step 4: Model Saving and Deployment

# Save the model
model.save('path/to/save/model')

# Step 5: Inference on New Data

# Load the saved model
model = tf.keras.models.load_model('path/to/saved/model')
```

```
# Inference on new data
for file in os.listdir('drive/MyDrive/toCheck'):
    test_image                              =
tf.keras.preprocessing.image.load_img('drive/MyDrive/toC
heck/' + file, target_size=(180, 180))
    test_image = image.img_to_array(test_image)
    test_image = np.expand_dims(test_image, axis=0)

    # Predictions
    predictions = model.predict(test_image)
    top_three_predictions                  =
get_top_three_predictions(predictions)

    # Store results
    resultDict[file] = top_three_predictions


    # Utility function to get top three predictions
    def get_top_three_predictions(predictions):
    class_labels = [i for i in range(1907, 1959)]
    class_labels.remove(1916)
    class_labels.remove(1917)
    class_labels.remove(1918)
    class_labels.remove(1920)

    # Convert predictions to list and find top three indices
    predictions_list = predictions.tolist()[0]
    sorted_predictions     =     sorted(predictions_list,
reverse=True)
    top_three_indices                      =
[predictions_list.index(sorted_predictions[i])   for   i   in
range(3)]

    # Map indices to class labels
    top_three_labels = [class_labels[idx]   for   idx   in
top_three_indices]
    return top_three_labels
```

5. Experiments

We have analyzed the set of undated manuscripts using both methods described in the paper, the manual signature method gave us the following results:

Table 1
Signature method

Period	Title
1910-1915	�-45
1910-1915	�-416
1910-1915	ꞯꞯ-650
1910-1915	ꞯꞯ-651-11
1910-1915	�-1330
1910-1915	�-1319
1910-1915	�-1920
1910-1915	�-1321
1910-1915	�-1322
1910-1915	�-1323
1910-1915	�-1324
1910-1915	�-1325
1910-1915	�-1326
1910-1915	�-1327
1910-1915	�-1361
1910-1915	�-2176
1910-1915	�-2176. 2
1916 -	ꞯꞯ-471-19
1916 -	ꞯꞯ-507-2
1916 -	ꞯꞯ-527-6
1916 -	ꞯꞯ-603-12
1916 -	ꞯꞯ-618-1
1916 -	ꞯꞯ-637-8
1916 -	ꞯꞯ-655-2
1916 -	�-1378-2
1916 -	Ꮟ-24551-242
1917-	ꞯꞯ-420
1917-	ꞯꞯ-488
1917-	ꞯꞯ-496
1917-	ꞯꞯ-638
1917-	ꞯꞯ-1678
1917-	ꞯꞯ-1718

The automated CNN method gave us the following results, it the period column there are three probable answers:

Table 2
Automotive CNN method

Period	Title
[1912, 1914, 1956]	�-45
[1910, 1926, 1930]	�-416
[1950, 1911, 1934]	ꞯꞯ-650
[1910, 1911, 1912]	ꞯꞯ-651-11
[1911, 1910, 1958]	�-1330
[1910, 1926, 1930]	�-1319
[1912, 1913, 1911]	�-1920
[1911, 1926, 1950]	�-1321
[1940, 1912, 1950]	�-1322
[1955, 1957, 1940]	�-1323
[1910, 1928, 1949]	�-1324
[1912, 1925, 1910]	�-1325
[1915, 1927, 1938]	�-1326
[1910, 1926, 1908]	�-1327
[1910, 1926, 1930]	�-1361
[1910, 1926, 1930]	�-2176
[1915, 1943, 1912]	�-2176. 2
[1925, 1926, 1922]	ꞯꞯ-471-19
[1958, 1950, 1914]	ꞯꞯ-507-2
[1925, 1926, 1926]	ꞯꞯ-527-6
[1908, 19406, 1950]	ꞯꞯ-603-12
[1925, 1926, 1908]	ꞯꞯ-618-1
[1922, 1925, 1950]	ꞯꞯ-637-8
[1949, 1926, 1925]	ꞯꞯ-655-2
[1950, 1955, 1910]	�-1378-2
[1940, 1955, 1908]	Ꮟ-24551-242
[1922, 1936, 1910]	ꞯꞯ-420
[1940, 1941, 1955]	ꞯꞯ-488
[1913, 1950, 1958]	ꞯꞯ-496
[1925, 1921, 1956]	ꞯꞯ-638
[1930, 1936, 1910]	ꞯꞯ-1678
[1908, 1909, 1910]	ꞯꞯ-1718
[1958, 1957, 1950]	3819

6. Conclusions

In evaluating manuscript dating, we compared results from the manual signature method and the automated Convolutional Neural Network (CNN) method, focusing on their alignment and discrepancies with probabilistic estimates from the CNN method. The manual signature method offers clear dating, and the CNN method aligns with this, showing high probabilities for certain dates while suggesting lower probabilities for others, indicating some uncertainty in the CNN model's accuracy. For broader ranges, the manual method's suggestions are somewhat supported by CNN's predictions with high and moderate probabilities, though some variability is highlighted. The manual method's datings align well with CNN's high-probability dates, though CNN's inclusion of low-probability dates suggests possible errors or wider uncertainty. Significant discrepancies arise when the manual method's datings diverge from CNN's predictions, indicating potential limitations or inaccuracies in the CNN model. For cases where the manual method's datings align closely with CNN's high probabilities, moderate and low probabilities show some acceptable variance but remain generally consistent. Both methods generally align for early manuscripts, with CNN's high-probability dates falling within the manual method's range, but significant differences are observed where CNN suggests dates outside the manual range, indicating potential issues with the CNN model's accuracy. The manual method provides consistent and reliable dating, while the CNN method introduces variability and highlights areas for further refinement to improve accuracy.

Acknowledgments

References

[1] K. Nesměrák, I. Němcová, Dating of Historical Manuscripts Using Spectrometric Methods: a mini-review, Analytical Letters 45(4) (2012) 330–344.

[2] E. Omayio, S. Indu, J. Panda, Historical Manuscript Dating: Traditional and Current Trends, Multimedia Tools and Applications 81(22) (2022) 31573–31602.

[3] L. MacKinney, Medical Illustrations in Medieval Manuscripts, Univ of California Press (2023).

[4] D. Antons, et al., The Application of Text Mining Methods in Innovation Research: Current State, Evolution Patterns, and Development Priorities, R&D Management 50(3) (2020) 329–351.

[5] A. Hamid, et al., Historical Manuscript Dating Using Textural Measures, International Conference on Frontiers of Information Technology (FIT), IEEE (2018).

[6] V. Dearing, Manual of Textual Analysis, Univ of California Press (2023).

[7] D. Van der Meij, Other Information on Dating and Ownership, Indonesian Manuscripts from the Islands of Java, Madura, Bali and Lombok, Brill (2017) 405–441.

[8] J. Droese, J. Karolewski, Manuscript Albums and Their Cultural Contexts: Collectors, Objects, and Practices, De Gruyter (2024).

[9] T. Tvalavadze, et al., Automated Dating of Galaktion Tabidze's Handwritten Texts, Advances in Computer Science for Engineering and Education VI, LNDECT 181 (2023). doi: 10.1007/978-3-031-36118-0_23.

[10] M. Iavich, M. Ninidze, Advancements in Dating Undated Manuscripts through Dual Methodologies, 29th International Conference "Information Society and University Studies" – IVUS 2024 (2024).

[11] G. Tabidze, Works in Fifteen Volumes 5 (2017) 250–251.

[12] F. Wahlberg, T. Wilkinson, A. Brun, Historical Manuscript Production Date Estimation Using Deep Convolutional Neural Networks, 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE (2016).

[13] A. Hamid, et al., Deep Learning Based Approach for Historical Manuscript Dating, International Conference on Document Analysis and Recognition (ICDAR), IEEE (2019).

[14] M. Boudraa, A. Bennour, Combination of Local Features and Deep Learning to Historical Manuscripts Dating, International Conference on Intelligent Systems and Pattern Recognition (2023).

[15] V. Yugay, et al., Stylistic Classification of Cuneiform Signs Using Convolutional Neural Networks, IT-Information Technology 0 (2024).

[16] A. Wang, et al., Repvit: Revisiting Mobile CNN from VIT Perspective, IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024).

[17] D. Bhatt, et al., CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope, Electronics 10(20) (2021).

[18] P. Verma, G. Foomani, Improvement in OCR Technologies in Postal Industry Using CNN-RNN Architecture: Literature review, Int. J. Machine Learning Comput. 12(5) (2022).

[19] V. Kharchenko, I. Chyrka, Detection of Airplanes on the Ground Using YOLO Neural Network, International Conference on Mathematical Methods in Electromagnetic Theory, MMET (2018) 294–297.