



DOI: 10.4312/mz.61.2.181-199

UDK: 781"17" Tartini G.(0.032)=131.1:004.352.242

A Digital Approach to Tartini's Theoretical Works: Training and Testing a Custom AI Model for Handwritten Text Recognition

Jerneja Umer Kljun

University of Ljubljana

ABSTRACT

The article describes the digitisation of Giuseppe Tartini's manuscripts with the help of handwritten text recognition via the Transkribus app. It explains the basics concepts, describes the training process for a customised model and the tests run on Tartini's scientific texts. The results of private and public AI models are compared in terms of accuracy, readability and efficiency of the editing process.

Keywords: Giuseppe Tartini, eighteenth century, HTR, Transkribus, digital humanities, digital musicology

IZVLEČEK

Članek prinaša podroben opis procesa digitalizacije rokopisov Giuseppeja Tartinija z uporabo programske opreme *Transkribus*, namenjene samodejnemu prepoznavanju in transkribiranju rokopisnega gradiva. V članku so predstavljena osnovna načela tehnologije HTR, opisana sta postopek učenja namenskega modela UI za transkribiranje ter preizkus dela z različnimi modeli pri transkribiranju Tartinijevih znanstvenih besedil. Sledi primerjava rezultatov, pridobljenih z dvema zasebnima in enim javno dostopnim modelom za transkripcijo italijanskih besedil, in sicer z vidika natančnosti, berljivosti in učinkovitosti, pri čemer je upoštevan tudi čas urejanja in revizije samodejno generirane transkripcije.

Ključne besede: Giuseppe Tartini, 18. stoletje, HTR, Transkribus, digitalna humanistika, digitalna muzikologija

* The author acknowledges the financial support from the Slovenian Research and Innovation Agency (research core funding no. P6-0446).

Introduction

The past decade has seen a renewed interest in Giuseppe Tartini, especially due to two consecutive Interreg project (tARTini¹ and TartiniBIS²) dedicated to the legacy of the composer as well as the preservation and promotion of his cultural heritage. However, even after the publication of Tartini's letters in 2020,³ and the launch of a digital collection⁴ and a thematic catalogue⁵ in 2025, much more remains to be uncovered.

The rapid advancements in artificial intelligence have created novel opportunities to access historical records. The article, meant as an account of an ongoing research project, presents a possible further step in the digitisation of Tartini's scientific texts kept at the Piran branch of the Regional Archives Koper (Pokrajinski arhiv Koper – Archivio regionale di Capodistria, hereinafter SI-PAK) through the integration of Handwritten Text Recognition (HTR) technology for automatic or assisted transcription provided by the Transkribus platform (READ Coop).⁶

Although an increasing number of researchers and institutions have begun exploring the use of HTR technology in various contexts,⁷ there is still room to further investigate its practical implementation and potential, as the field is constantly evolving. This article contributes to that growing body of scholarly work by providing a detailed account of the HTR software used, namely the Transkribus platform, including the training process for a tailored text recognition model, i.e. a trained AI algorithm, able to detect the most probable sequence of characters for each text line,⁸ its advantages and limitations, and its possible applications in future interdisciplinary research and teaching, especially in the field of digital musicology.

To this end, the article first gives a brief introduction to the general concepts of HTR and describes the main features of the Transkribus app. It then focuses

1 "tARTini", *Interreg Italia – Slovenija*, accessed September 5, 2025, <https://2014-2020.ita-slo.eu/it/tartini>.

2 "TartiniBIS", *Interreg Italia – Slovenija*, accessed September 5, 2025, <https://www.ita-slo.eu/it/tartini-bis>.

3 Giorgia Malagò (ed.), *Giuseppe Tartini / Pisma in Dokumenti / Letters and Documents*, vol. 1, transl. Jerneja Umer Kljun and Roberto Baldo (Trieste: EUT Edizioni Università di Trieste, 2020).

4 "Archives and Catalogue", *Discover Tartini*, accessed September 5, 2025, <https://www.discovertartini.eu/archivi?type=1&lang=en>.

5 "Thematic Catalogue", *Discover Tartini*, accessed September 5, 2025, <https://www.discovertartini.eu/archivi?type=5&lang=en>.

6 "Transkribus", *Transkribus*, accessed August 20, 2025, <https://www.transkribus.org/>.

7 Joe Nockels et al., "Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research", *Archival Science* 22, no. 3 (2022): 367–392.

8 "Training Text Recognition Model", *Transkribus*, accessed September 5, 2025, <https://help.transkribus.org/training-text-recognition-models>.

on the archival sources selected for a test run with Transkribus, describing the main features of Tartini's theoretical manuscripts. In addition, an experiment was conducted to test the effectiveness of the app by comparing the results of automatic transcription with two custom AI models with the results of a publicly available AI model for handwritten Italian, in terms of readability, the ability of the model to accurately recognize numbers, abbreviations, ratios and fractions, as well as the estimated time needed for editing a single automatically recognized page.

A Brief Introduction to Handwritten Text Recognition

Handwritten text recognition, i.e. the technology that generates machine-readable transcriptions from images of manuscripts, is certainly not a novelty – according to Muehlberger et al., it is “an active research area in the computational sciences dating back to the mid-twentieth century”.⁹ Originally closely linked to the development of optical character recognition (OCR), which is used to convert scanned images of printed text into machine-encoded text, it quickly became “a research area in its own right”¹⁰ due to the computational complexity involved in dealing with the specificity of handwriting, with its frequent deviations from the norm, as well as the variability of different handwriting styles.¹¹ And while the OCR problem has been solved rather early and effectively, as there is (only) a finite number of fonts that can be used in print, which allows the software to be programmed to recognize and read them all, the nearly infinite number of possibilities rendered the same task impossible when dealing with handwriting.¹²

The HTR technology made considerable leaps forward with the advent and improvement of artificial neural networks.¹³ With the use of AI models and Deep Learning, HTR software is now able to *learn to recognize* handwriting in an image, using “the model's knowledge to transcribe that handwriting into digital text”.¹⁴ But as the topic of deep neural network approaches exceeds the scope of this article, readers should refer to the cited sources as well as the numerous user-friendly resources available on the READ-COOP platform

9 Guenter Muehlberger et al., “Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study”, *Journal of Documentation* 75, no. 5 (2019): 995.

10 Ibid., 955.

11 Ibid.; Jože Glavič, “Primer uporabe programa Transkribus in izdelava modela za avtomatsko optično prepoznavanje znakov za poenostavljeno transkribiranje ročno pisane gotice”, *Moderna arhivistika* 3, no. 1 (2020): 87.

12 “What Is HTR and How Does It Work”, *Transkribus blog*, accessed September 5, 2025, <https://blog.transkribus.org/en/what-is-handwriting-recognition-and-how-does-it-work>.

13 Ibid., 79.

14 “What Is HTR and How Does It Work”.

for a detailed overview of the technology behind Transkribus.¹⁵ From a user's perspective, it is sufficient to note that HTR: "is now a mature machine learning tool capable of producing accurate, machine-processable text from images of historical manuscripts [...] speeding up the transcription of primary sources and facilitating full text searching and analysis of historic texts at scale."¹⁶

Working with Transkribus

Transkribus is a comprehensive AI-powered, user-friendly HTR software, developed within the framework of the Recognition and Enrichment of Archival Documents (READ) European Union Horizon 2020 project, which is now run by the READ-COOP cooperative.¹⁷ The Transkribus app allows users to quickly transcribe handwritten documents using one of the many publicly available text recognition models (at the moment of writing this article, there are 325 public models trained on a variety of scripts and languages). However, it is much more effective if it is trained to recognise a specific hand. The workflow for training a custom text recognition model in Transkribus consists of several steps, which can be roughly summarised as follows:

- i. Layout recognition (segmentation):
after uploading an image of a manuscript, the automatic layout recognition function segments the image into text regions and lines that connect the text and the image and serve as a reference point for text recognition;
- ii. manual transcription or text-image matching and editing (data preparation):
the editing software shows the layout editor and the text editor simultaneously, allowing the user to manually transcribe the page line by line. The transcription can also be enriched by using textual tags to mark-up words (e.g. names, dates, places) or to add other information about the manuscript. Once completed, the transcription can be downloaded in different formats (image, searchable PDF, docx, text file and XML). It is now also possible to align pre-existing transcriptions with images of manuscripts;¹⁸
- iii. training:
accurately transcribed pages serve as *ground truth*, i.e. accurate and verified data used for training and testing AI models.¹⁹ To train a solid HTR model, it is advisable to manually transcribe at least 10,000 words for each

15 *Transkribus Blog*, accessed September 5, 2025, <https://blog.transkribus.org/en>.

16 Nockels et al., "Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts", 368.

17 Muehlberger et al., "Transforming Scholarship in the Archives through Handwritten Text Recognition".

18 "Text-Image Matching", *Transkribus*, accessed September 5, 2025, <https://help.transkribus.org/import-existing-transcriptions-with-text-image-matching>.

19 "What Is Ground Truth?", *Transkribus*, accessed September, 2025, <https://help.transkribus.org/import-existing-transcriptions-with-text-image-matching>.

hand. After choosing the training data, i.e. ground truth data on which the knowledge of the model is based, and the validation data, i.e. a set of examples from the same collection that serves as evaluation of the model, the users can proceed to set up and train their model.²⁰ During the training process, the validation data is set aside to serve to test the model's performance. By comparing the automatically transcribed pages against the accurate manually transcribed data, the model counts all the incorrectly recognized characters, giving a "score" to the overall performance. The metric used is the *character error rate* or CER, i.e. the number of errors expressed as a percentage of the total number of transcribed characters.²¹ This exact process has been applied to a selection of documents described in the next section to train a custom model for the transcription of Tartini's theoretical work.

Towards a Digitised Tartini Collection

The Tartini family collection is succinctly described in Pucer's introduction to the catalogue:²²

[The Tartini collections] *was previously stored at the "Sergej Mašera" Maritime Museum in Piran, where some of the documents were also displayed as exhibits. The collection was organised in a makeshift and non-archival manner. The documents were numbered; however, they were not organised chronologically or by subject. An inventory of sorts was included in the form of a manuscript written by Igor Cvetko. In 1986, the collection was acquired by the Piran branch of PAK.*²³

Pucer started working on the collection in 1991, compiling a detailed inventory of the material. The collection, covering the period from 1654 to 1951, comprises 11 archival units, i.e. 1.1 linear meter, organised in seven thematic sections, indexed with the letters A through G, namely A – Family records; B – Scientific texts; C – Copies of Tartini's work; D – Commemoration of the 200th anniversary of Tartini's birth in Piran; E – Print; F – Photography;

20 "Setup and Training", *Transkribus Blog*, accessed September, 2025, <https://help.transkribus.org/model-setup-and-training>.

21 "How Is CER Calculated in Transkribus?", *Transkribus Blog*, accessed September 5, 2025, <https://blog.transkribus.org/en/how-is-the-cer-calculated-in-transkribus>.

22 Albert Pucer, *Giuseppe Tartini: Inventar zbirke / Inventario della collezione: 1654–1951* (Koper: Pokrajinski arhiv Koper, 1993).

23 "Arhivska zbirka 'Giuseppe Tartini' je bila prej shranjena v Pomorskem muzeju 'Sergej Mašera' v Piranu, kjer so nekateri dokumenti služili tudi kot razstavnii eksponati. Tu je bila zbirka zasilno in nearhivsko urejena. Dokumenti so bili oštevilčeni, vendar niso bili urejeni ne po zadevah ne kronološko. V rokopisu je bil izdelan tudi neke vrste inventar, delo mag. Igorja Cvetka. Leta 1986 je zbirko dobil PAK – enota v Piranu." Pucer, *Giuseppe Tartini*, 7, quotation translated by Jerneja Umer Kljun.

G – Miscellanea. Each section is further divided into subsections marked with roman numerals (I – XIV) and the documents are organised chronologically.²⁴

Within the framework of the TartiniBIS project, Pucer's catalogue served as the basis for the identification of previously undiscovered or undigitised archival records. A comprehensive examination of the digital and physical archive uncovered various problems, such as difficult access, incomplete digitisation, digitisation errors and missing or misplaced documents.²⁵ And while the missing materials have now been digitised and published on the discovertartini.eu website,²⁶ the PDF documents and images available online are still scattered across various databases and out of context. To address the issue of document accessibility, further efforts could also be made by applying the methods presented here.

Description of the Digitally Transcribed Manuscripts

The small selection of seven digitally transcribed documents from the “scientific” section of the Tartini fonds presented in this article is built upon a manual transcription of Tartini's unpublished treatise *Quadratura del circolo* (see entry #1 in Table 1 below), previously discussed by Sukljan²⁷ and Umer Kljun²⁸ in the volume *In Search of Perfect Harmony: Tartini's Music and Music Theory in Local and European Contexts*.²⁹ The manual transcription of the relatively short treatise proved to be optimal as ground truth, as it yielded the first successful version of the Giuseppe Tartini model, with a CER of 2.3%. Six more short unpublished documents that have not been transcribed before were then chosen from the Tartini fonds, t. u. 3 and 4 (scientific texts), automatically transcribed and manually edited to serve as ground truth for further model training and testing.

24 Ibid., 7.

25 “Five Studies and New Sources”, *Discover Tartini*, accessed September 5, 2025, <https://www.discovertartini.eu/main/pagina/19/Five-studies-and-new-sources?lang=en>.

26 “Nuove fonti dagli archivi di Pirano”, *Discover Tartini*, accessed September 5, 2025, <https://www.discovertartini.eu/index.php/archivi/pirano>.

27 Nejc Sukljan, “Tartini and the Ancients: Traces of Ancient Music Theory in the Tartini-Martini Correspondence”, in *In Search of Perfect Harmony: Tartini's Music and Music Theory in Local and European Contexts*, ed. Nejc Sukljan (Berlin: P. Lang, 2022), 141–167.

28 Jerneja Umer Kljun, “Understanding Tartini and His Thought: Overcoming Translation Difficulties in the Correspondence between Tartini and Martini”, in *In Search of Perfect Harmony: Tartini's Music and Music Theory in Local and European Contexts*, ed. Nejc Sukljan (Berlin: P. Lang, 2022), 245–260.

29 Nejc Sukljan (ed.), *In Search of Perfect Harmony: Tartini's Music and Music Theory in Local and European Contexts* (Berlin: P. Lang, 2022).

Table 1: The Digitised HTR Tartini Collection

Entry #	Identifier	# Pages Transcribed	Working title	Opening lines	Genre	Contents, themes
1	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 4, a. u. 232	52	<i>Quadratura del circolo</i>	“Si è scoperto un fenomeno armonico, per di cui mezzo si pretende ottenuta la quadratura del circolo [...]”	scientific text	third tone; squaring of the circle
2	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 3, a. u. 143	7	<i>Risposta a critico ignoto</i>	“Monsieur, Se la fatica di tanti anni per la ricerca del vero principio dell'armonia non mi avesse recato altro vantaggio [...]”	scientific text, letter	third tone; a defence of the consequences deduced from the third tone
3	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 3, a. u. 150	8	<i>Proseguimento e compimento della Carta Musicale. La prima Carta musicale</i>	“Proseguimento, e compimento della Carta Musicale. La prima proposizione fu l'accordo consonante simultaneo della sestupla armonica [...]”	didactic material	music theory; a paper on ratios, progressions, intervals and notation
4	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 3, a. u. 151	30	<i>Dissertazione musicale sui principi dell'armonia</i>	“L'Autore, che nel suo Trattato di Musica secondo la vera Scienza dell'armonia [...]”	scientific text	the true foundation of harmony
5	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 3, a. u. 153	24	<i>Parte Terza</i>	“In questa terza parte s'intende dimostrare, che la scienza del presente sistema sia identica: ⁶ quella stessa de Pittagorici e di Platone [...]”	scientific text	a justification of the proposed harmonic system through an elaboration of Plato's Timaeus

Entry #	Identifier	# Pages Transcribed	Working title	Opening lines	Genre	Contents, themes
6	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 3, a. u. 156	12	<i>Evidentissima dimostrazione</i>	“Evidentissima dimostrazione dell’armonica natura di quantità <u>a</u> priori.”	scientific text	the intrinsic harmonic nature of the One as a whole
7	SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 4, a. u. 241	3	<i>Confronto dei principj</i>	“Confronto dei principj nel piano dimostrativo. Principj ipotetici: punto, linea, e superficie: delle Scienze Comuni [...]”	scientific text	a comparison of the fundamental principles of the commonly known sciences and the principles of Tartini’s harmonic science

All seven autographs, written in eighteenth century cursive, present similar traits and while Tartini's penmanship is quite legible and consistent, there are some orthographic variations that could prove useful when placing the documents on a timeline or when trying to connect various scattered fragments. For example, there is some oscillation in the use of double consonants, such as -z- and -zz- in *mezo*, *mezi* and *mezzo*, *mezzi* (mean, means).

Furthermore, "Parte terza"³⁰ includes segments in Latin, in which Tartini employs the *e caudata* [ɛ] for *æ* (see Figure 1), and three of the seven digitised documents present an uncommon diacritic above the letter *e* that does not appear elsewhere and the (inconsistent) use of which remains uncertain (see underlined words in Figure 1 and excerpts from various documents in Figure 2).

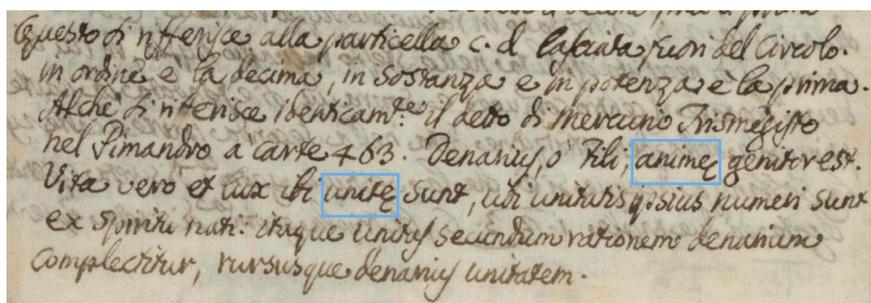


Figure 1: Use of the e caudata in Latin script.

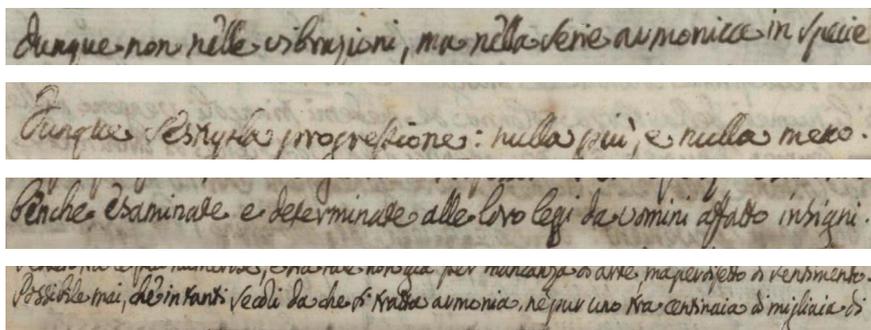


Figure 2: Uncommon diacritics in Tartini's manuscript.

In Tartini's texts, the most frequently abbreviated words are Italian adverbs of manner ending in *-mente* (e.g. *demostrativamente*, *universalmente*) and qualitative adjectives describing roots, ratios and progressions (e.g. *armonico*,

30 SI PAK PI 334 Zbirka Giuseppe Tartini, t. u. 3, a. u. 153 ("Parte Terza").

aritmetico, geometrico) or superlatives (*verissimo*), all categories abbreviated with the use of a colon and superscript (e.g. *demonstrativam.^{te}*, *geom.^{che}*, *ver.^{mo}*). Other methods of abbreviation include the use of a macron over an omitted letter or a longer stroke indicating contractions or suspensions, e.g. *prop̄te* for *proporzionale*, *caplo* for *capitolo*, as well as various forms of ordinal indicators, e.g. *2dō*, *3zā* and even *sesqui3zā*, as shown in Figure 3 below.

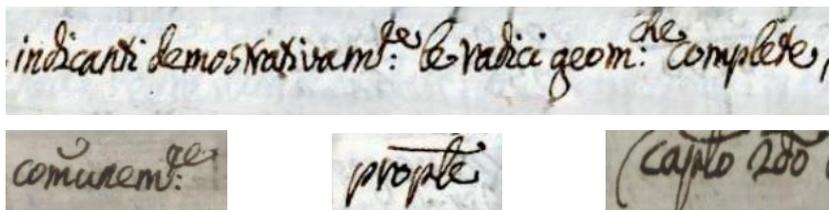


Figure 3: Frequently used abbreviations.

In general, the *mise-en-page* is relatively straightforward, facilitating rapid automatic layout recognition. However, most of the digitised theoretical texts are characterized by the presence of specific graphical elements, such as musical and mathematical symbols that appear both in-line, which means they are embedded in the sentence structure and therefore must be transcribed for the sentence to be complete, as well as separate, floating elements,³¹ see Figure 4 below. Similarly, Tartini makes frequent use of illustrative examples of both musical and mathematical nature, i.e. geometric drawings, sketches, calculations and musical examples, see Figure 5 below.

Since the primary goal was to obtain an effective model, a “hyper-diplomatic” approach to transcription proved to be the best option in the data preparation stage, “so as not to confuse the computer model when it is comparing transcriptions to the image of the manuscript”.³² This means that the manual transcription follows the source material as closely as possible, including non-standard punctuation and capitalization, superscript, un-expanded abbreviations etc. However, due to the specificities of Tartini’s text some adjustments and simplifications had to be made (e.g. omission of in-line graphics when a suitable Unicode character was not available).

31 Fabian C. Moss et al., “Digitizing a 19th-Century Music Theory Debate for Computational Analysis”, in *Proceedings of the Conference on Computational Humanities Research 2021: Amsterdam, the Netherlands, November 17–19, 2021*. https://ceur-ws.org/Vol-2989/short_paper31.pdf.

32 Bram Caers, “Teaching Handwritten Text Recognition: Can New Technologies Save Old Skills?”, *Quaerendo* 54, nos. 2-3 (2024): 207.

Testing Model Versions

To test the two versions of the model specifically trained on Tartini's handwriting, confronting the fully automatic transcription to the one achieved with a publicly available model for Italian, a single page from the Tartini archive, which has not been previously transcribed, has been selected. To ensure the best possible outcome with each of the selected AI models, the test page had to present minimal "disturbance", i.e. clear handwriting, little to no cancellations, corrections, additions, ink bleed-through or lacerations. It also had to include abbreviations, numbers, ratios, fractions and calculations, as they are consistently present in Tartini's theoretical work. The effectiveness of the automated transcription was then assessed in terms of readability, the ability of the model to correctly transcribe abbreviations, ratios and fractions, as well as the estimated time needed for editing a single page.

The models used to transcribe the selected page from "Appunti sulla proporzione armonica"³³ were *Transkribus Italian handwriting M1*³⁴ (trained by the Transkribus team on 653,630 words; 6.70% CER) and my own private models Giuseppe Tartini I (trained on 26,402 words; 2.30% CER) and Giuseppe Tartini II (trained on 51,355 words; 3.06% CER). *The Italian handwriting M1* is a generic model trained on a diverse dataset spanning from the sixteenth to the nineteenth century and it is occasionally updated with community data. *Giuseppe Tartini I* is a private model trained on the manual transcription of Tartini's unpublished manuscript "Quadratura del circolo" (diplomatic transcription; abbreviations tagged; expansions provided within the tag). *Giuseppe Tartini II* is a private model built upon the previous version of the model. It is trained on the Tartini Collection: one manual transcription of Tartini's unpublished manuscript "Quadratura del circolo"; six documents, automatically recognized with Giuseppe Tartini I model, which were then edited and tagged (diplomatic transcription; abbreviations tagged; expansions provided within the tag; documents feature tables and special Unicode characters; training included expanding abbreviations).

The differences between the resulting transcriptions are best illustrated by the examples in Tables 2 and 3 below, in which errors are highlighted.

33 Biblioteca del Conservatorio di musica Giuseppe Tartini, IT-TS0108, Collezione Tartiniana: scritti teorici, musiche e documenti relativi a Giuseppe Tartini in Trieste e Pirano: GT/FA/23 ms 1, "Appunti sulla proporzione armonica".

34 "Transkribus Italian Handwriting M1", *READ Coop*, accessed September 5, 2025, <https://app.transkribus.org/models/text/38440>.

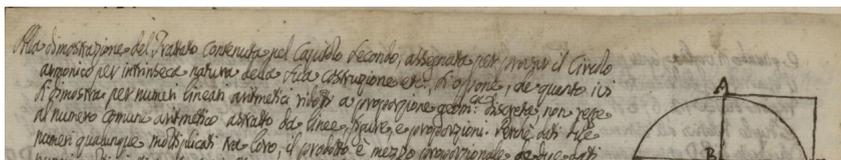


Figure 6: Excerpt from “Appunti sulla proporzione armonica” (GT/FA/23 ms1), top of the page.

Table 2: Unedited automatic transcriptions with selected models

Public model	Alla dimostrazione del trattato contenuta nel Capuoto secondo, assegnata per sopra il Circolo armonico per intrinseca natura della sua costruzione et soglione, che quanto i si dimostra per numesi linean animatici ridotto a proporzione alm.te di spese, non rese al numero mune a rimetto attratto da linee, pare, e predazioni. Perche dati dice paesi qualunque moltiplicati tra loro, il prodotto è mezzo proporzionale de due da
Giuseppe Tartini I	Alla dimostrazione del trattato contenusa nel circolo secondo assegnata per provati il circolo armonico per intrinseca natura, della sula costragione etc.; si ossone, che quanto ivi si dimostra per numeri lineari aritmetici, ridotti a proporzione geom:ca discreta, non regge al numero comune aritmetico astratto da linee, figure, e proporzioni. Perchè dati due numeri qualunque moltiplicati tra loro, il prodotto è mezzo proporzionale de due dati
Giuseppe Tartini II	Alla dimostrazione del Trattato contenuta nel capitolo secondo assegnata per provare il circolo armonico per intrinseca natura della sua costruzione etc.; si ossione, che quanto ivi Si dimostra per numeri linea aritmetici ridotti a proporzione geom:ca discreta, non regge al numero comune aritmetico astratto da linee, figure, e proporzioni. Perchè dati due numeri qualunque moltiplicati tra loro, il prodotto è mezzo proporzionale de due dati

Similarly, if we observe a section with a more complicated layout, the efficacy of the models becomes even more obvious:

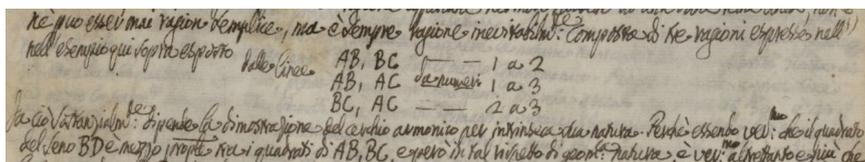


Figure 7: Excerpt from “Appunti sulla proporzione armonica” (GT/FA/23 ms1), bottom of the page.

Table 3: Unedited automatic transcriptions with selected models

Public model	<p>ne puo esser mai ragion semplice, ma è sempre ragione, inevitabilm.te composte di tre ragioni espresse nell' nell'esempio, qui sopra et poi</p> <p>488 142 dalle lino A. A da vera 3 846 243</p> <p>la citta siam.te dipende la dimostrazione del cerchio armonico per intrinseca la natura esse essendo vero de il quadra del seno [BD] è mezzo [prople] non tra i quadrati di 8, e però in tal rispetto di pom.a natura, e ver altrettanto e più che</p>
Giuseppe Tartini I	<p>ne può esser mai ragion semplice, ma è sempre ragione inevitabilm:te composta di tre ragioni espresso nelle nell'essenqui qui sopra esposta</p> <p>AB, BC, 1 a 2, dalle linee, Ab, Ac da umeri 1 a 3 BC, AC, = 2 a 3</p> <p>Sa ciò sottavialm:te dipende la dimostrazione del cerchio armonico per intrinseca sia natura. Perchè essendo verte che il quadrato del seno BD è mezzo prople tra quadrati di AB, BC, e però in tal rispetto di geom:ca natura, è veri altrettanto, e sia, che</p>
Giuseppe Tartini II	<p>nè puo esser mai ragion semplice, ma è dempre ragione inevitabilm:te composta di tre ragioni espresse nell' nell'esempio qui sopra esposta</p> <p>AB, BC 1 a 2 dalle linee AB, AC, da numeri 1 a 3 BC, AC 2 a 3</p> <p>Da ciò sostanzialm:te dipende la dimostrazione del cerchio armonico per intrinseca sua natura. Perchè essendo ved:mo che il quadrato, del seno BD è mezzo prople tra i quadrati di AB, BC, e però in tal rispetto di geom:ca natura, è veri altrettanto e più, che</p>

Discussion

In general, it can be observed that using a public model when automatically transcribing Tartini's script is not really a time-saving solution. As expected, the model used in this comparison proved to be unsuitable for the task, as it had difficulties recognizing abbreviations and even differentiating letters and numbers, which yielded a nearly illegible, nonsensical transcription that would possibly take longer to edit than transcribing the document manually. The two versions of the private model give far better results, even though several corrections are still necessary. While the text is readable and errors mainly occur at character level (e.g. interchanging the letters q, g, z, p and r, f and s), a bigger issue stems from an inaccurate layout recognition and line segmentation (see Table 3 above) – an issue that can be easily prevented with proper layout editing before proceeding to text recognition.

To quantify these observations, each of the automated transcriptions was manually corrected and edited, the editing process was timed, and the version history in Transkribus and the comparison function in Word were used to identify the changes in the resulting transcriptions. While it only took a little over a minute to automatically transcribe a page with each of the models, it took over 47 minutes to correct the layout and the text transcribed with the public AI model. A comparison of the exported text files transcribed with Italian Handwriting M1 showed that 370 revisions were made to the text itself, while the Transkribus version history feature calculated a 20.79% CER and a 57.52 % WER (word error rate) on the analysed page. In comparison, it took 25 minutes to complete the same task on the page transcribed using the Giuseppe Tartini I model (199 text revisions detected by document comparison; 5.72% CER and 15.22% WER calculated with 'Transkribus' version history feature), and 19 minutes to correct the layout and text transcribed with the AI Giuseppe Tartini II model (98 revisions detected by document comparison, 3.47% CER and 8.48% WER calculated with 'Transkribus' version history feature). It should be noted that half of that time was spent on layout editing, which means the text editing will be quicker if the layout of the pages with a more complex structure is carefully prepared in advance.

To summarise, the performance of the second version of the private model is already quite satisfactory, but it could be improved further with more data and "fine-tuning", as there are still some issues to be addressed. For example, most floating elements had to be excluded from the training process because the software cannot yet recognise musical notation or other graphics. However, efforts have been made towards handwritten music recognition (OMR) within Transkribus, which entails textual encoding of the written music.³⁵ Whi-

35 "Jorge Calvo-Zaragoza – Handwritten Music and Text Recognition in Transkribus", YouTube video, 08:02, accessed September 5, 2025, <https://www.youtube.com/watch?v=TvDyF9L7sYE>.

le the method proposed by Calvo-Zaragoza³⁶ did not seem efficient enough at this time or even compatible with training an HTR model, the automatic recognition of Tartini's music may be well within reach in the near future.

In Lieu of a Conclusion: Possible Applications

As the digitisation of the Tartini collection is an ongoing project, there is no concrete conclusion to be made yet. It is easy to see, however, how this approach could be useful in compiling a fully digitised, machine-readable database of Tartini's scholarly work that could be adapted for research (e.g. corpus-based studies) or critical editions. The process of model training is a time-consuming process, nevertheless, the efficiency of the tool is truly expressed when working with a large collection of documents,³⁷ creating a large data set. And as Nicholas Cook observed, "working with larger data sets will open up new areas of musicology".³⁸ In addition, the use of HTR in university courses has already proven to be an engaging method for teaching palaeography and digital editing, revitalizing the discussion about text editions.³⁹ With HTR (and AI in general), we are standing at a crucial moment in digital humanities, a moment that requires a sensible and well-considered approach, but nevertheless "a moment of opportunity"⁴⁰.

Bibliography

Primary sources

Biblioteca del Conservatorio di musica Giuseppe Tartini, IT-TS0108

Collezione Tartiniana: scritti teorici, musiche e documenti relativi a Giuseppe Tartini in Trieste e Pirano

GT/FA/23 ms1 ("Appunti sulla proporzione")

Pokrajinski arhiv Koper/Koper Regional Archives

SI PAK PI 334 Zbirka Giuseppe Tartini

t. u. 3

a. u. 143 ("Risposta a critico ignoto")

a. u. 150 ("Proseguimento e Compimento della Carta musicale")

a. u. 151 ("Dissertazione musicale sui principi dell'armonia")

a. u. 153 ("Parte Terza")

a. u. 156 ("Evidentissima dimostrazione")

t. u. 4

a. u. 232 ("Quadratura del circolo")

a. u. 241 ("Confronto dei principj")

36 Ibid.

37 Glavič, "Primer uporabe programa transkribus [...]", 88.

38 Nicholas Cook, "Towards the Complete Musicologist?", *ISMIR 2005*, 4, <https://ismir2005.ismir.net/documents/Cook-CompleatMusicologist.pdf>.

39 Caers, "Teaching Handwritten Text Recognition".

40 Cook, "Towards the Complete Musicologist?", 1.

References

- “Archives and Catalogue.” *Discover Tartini*. Accessed September 5, 2025. <https://www.discovertartini.eu/archivi?type=1&lang=en>.
- Caers, Bram. “Teaching Handwritten Text Recognition: Can New Technologies Save Old Skills?” *Quaerendo* 54, nos. 2-3 (2024): 198–209. https://brill.com/view/journals/qua/54/2-3/article-p198_6.pdf; DOI: 10.1163/15700690-BJA10024.
- Cook, Nicholas. “Towards the Complete Musicologist?.” *ISMIR 2005*. <https://ismir2005.ismir.net/documents/Cook-CompleatMusicologist.pdf>.
- “Five Studies and New Sources of Documentation Relating to Giuseppe Tartini, His Contemporaries, Pupils and Followers.” *Discover Tartini*. Accessed September 5, 2025. <https://www.discovertartini.eu/main/pagina/19/Five-studies-and-new-sources?lang=en>.
- Glavič, Jože. “Primer uporabe programa transkribus in izdelava modela za avtomatsko optično prepoznavanje znakov za poenostavljeno transkribiranje ročno pisane gotice.” *Moderna arhivistika* 3, no. 1 (2020): 86–97. DOI: <https://doi.org/10.54356/MA/2020/1/ECAD5603>.
- “How Is CER Calculated in Transkribus.” *Transkribus Blog*. Accessed September 5, 2025. <https://blog.transkribus.org/en/how-is-the-cer-calculated-in-transkribus>.
- “Jorge Calvo-Zaragoza – Handwritten Music and Text Recognition in Transkribus.” YouTube video. 08:02. Posted by Transkribus on 12 october 2022. Accessed September 5, 2025. <https://www.youtube.com/watch?v=TvDyF9L7sYE>.
- Malagò, Giorgia, ed. *Giuseppe Tartini / Pisma in Dokumenti / Letters and Documents*. Vol. 1. Translated by Jerneja Umer Kljun and Roberto Baldo. Trieste: EUT Edizioni Università di Trieste, 2020.
- Moss, Fabian C., Maik Köster, Melinda Femminis, Coline Métrailler, and François Bavaud. “Digitizing a 19th-Century Music Theory Debate for Computational Analysis.” In *Proceedings of the Conference on Computational Humanities Research 2021: Amsterdam, the Netherlands, November 17–19, 2021*. https://ceur-ws.org/Vol-2989/short_paper31.pdf.
- Muehlberger, Guenter, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, et al. “Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study.” *Journal of Documentation* 75, no. 5 (2019): 954–976. <https://doi.org/10.1108/JD-07-2018-0114>.
- Nockels, Joe, Paul Gooding, Sarah Ames, and Melissa Terras. “Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research.” *Archival Science* 22, no. 3 (2022): 367–392. <https://doi.org/10.1007/s10502-022-09397-0>.
- “Nuove fonti dagli archivi di Pirano.” *Discover Tartini*. Accessed September 5, 2025. <https://www.discovertartini.eu/index.php/archivi/pirano>.
- “OCR vs. HTR.” *Transkribus Blog*. Accessed September 5, 2025. <https://blog.transkribus.org/en/insights/ocr-vs-htr>.
- Pucer, Albert. *Giuseppe Tartini: Inventar Zbirke / Inventario Della Collezione: 1654–1951*. Koper: Pokrajinski arhiv Koper, 1993.
- “Setup and Training.” *Transkribus Blog*. Accessed September 5, 2025. <https://help.transkribus.org/model-setup-and-training>.
- Sukljan, Nejc, ed. *In Search of Perfect Harmony: Tartini's Music and Music Theory in Local and European Contexts*. Berlin: P. Lang, 2022. DOI: 10.3726/b20325.
- Sukljan, Nejc. “Tartini and the Ancients: Traces of Ancient Music Theory in the Tartini-Martini Correspondence.” In *In Search of Perfect Harmony: Tartini's Music and Music*

- Theory in Local and European Contexts*, edited by Nejc Sukljan, 141–167. Berlin: P. Lang, 2022.
- “TartiniBIS.” *Interreg Italia – Slovenija*. Accessed September 5, 2025. <https://www.ita-slo.eu/it/tartini-bis>.
- “TARTini.” *Interreg Italia – Slovenija*. Accessed September 5, 2025. <https://2014-2020.ita-slo.eu/it/tartini>.
- “Text-Image Matching.” *Transkribus*. Accessed September 5, 2025. <https://help.transkribus.org/import-existing-transcriptions-with-text-image-matching>.
- “Thematic Catalogue.” *Discover Tartini*. Accessed September 5, 2025. <https://www.discovertartini.eu/archivi?type=5&lang=en>.
- “Training Text Recognition Model.” *Transkribus*. Accessed September 5, 2025. <https://help.transkribus.org/training-text-recognition-models>.
- Transkribus Blog*. Accessed September 5, 2025. <https://blog.transkribus.org/en>.
- “Transkribus Italian Handwriting M1.” *READ Coop*. Accessed September 5, 2025. <https://app.transkribus.org/models/text/38440>.
- “Transkribus.” *Transkribus*. Accessed August 20, 2025. <https://www.transkribus.org/>.
- Umer Kljun, Jerneja. “Understanding Tartini and His Thought: Overcoming Translation Difficulties in the Correspondence between Tartini and Martini.” In *In Search of Perfect Harmony: Tartini’s Music and Music Theory in Local and European Contexts*, edited by Nejc Sukljan, 245–260. Berlin: P. Lang, 2022.
- “What Is Ground Truth?” *Transkribus*. Accessed September 5, 2025. <https://help.transkribus.org/import-existing-transcriptions-with-text-image-matching>.
- “What Is HTR and How Does It Work.” *Transkribus Blog*. Accessed September 5, 2025. <https://blog.transkribus.org/en/what-is-handwriting-recognition-and-how-does-it-work>.

SUMMARY

The article outlines the potential next steps in digitising Giuseppe Tartini’s scientific texts kept at the Piran branch of the Regional Archives Koper, through the integration of Handwritten Text Recognition (HTR) technology for automatic or assisted transcription. It first gives a brief overview of the general concepts of HTR, describing the main features of the Transkribus app, including the training process for a customized text recognition model. The focus then shifts towards the primary sources selected for a test run with the HTR app, describing the main features of Tartini’s theoretical manuscripts. In addition, an experiment was conducted to test the performance of the trained model by comparing the results of two private custom AI models against the results obtained with a publicly available AI model for handwritten Italian. The resulting transcriptions were compared in terms of readability, the ability of the model to accurately recognize numbers, abbreviations, ratios and fractions, as well as the estimated time needed for editing a single automatically recognized page.

POVZETEK

Digitalni pristop pri obravnavi Tartinijevih teoretičnih del: učenje in preizkušanje prilagojenega umetnointeligentnega modela za prepoznavanje rokopisnih besedil

Članek zarisuje eno izmed možnih nadaljnjih poti pri digitalizaciji znanstvenih besedil Giuseppeja Tartinija, ki jih hrani piranska enota Pokrajinskega arhiva Koper, in sicer z uporabo tehnologije HTR (*handwritten text recognition*), namenjene samodejnemu prepoznavanju in transkribiranju rokopisnega gradiva. Uvod prinaša kratek pregled osnovnih načel tehnologije HTR in opisuje glavne značilnosti programske opreme Transkribus, ki omogoča učenje prilagojenih umetnointeligentnih modelov za prepoznavanje rokopisnih besedil. Po podrobni razčlenitvi temeljnih značilnosti izbranih Tartinijevih rokopisov, tj. primarnih virov, izbranih za preizkušanje programske opreme, sledi opis preizkusa učinkovitosti izbranega pristopa za obdelavo gradiva. Nazadnje je predstavljena primerjava rezultatov dveh zasebnih prilagojenih umetnointeligentnih modelov z rezultati, pridobljenimi z javno dostopnim modelom za italijanščino, in sicer z vidika natančnosti pri prepoznavanju in prepisovanju števil, krajšav, razmerij in ulomkov, berljivost besedila in splošni učinkovitosti modelov, pri čemer je upoštevan tudi čas urejanja in revizije samodejno generirane transkripcije.

ABOUT THE AUTHOR

JERNEJA UMER KLJUN (Jerneja.umerkljun@ff.uni-lj.si) is a translator, translation teacher and researcher exploring the fields of Translation Studies, Sociolinguistics and Digital Humanities. She teaches Italian-Slovene translation and several language courses at the Department of Translation Studies (Faculty of Arts, University of Ljubljana) as well as at the Academy of Music of the University of Ljubljana. She has translated Giuseppe Tartini's letters into Slovene and authored several scientific articles and a monograph on code-switching and language attitudes.

O AVTORICI

JERNEJA UMER KLJUN (Jerneja.umerkljun@ff.uni-lj.si) je prevajalka, visokošolska učiteljica in raziskovalka na področju prevodoslovja, sociolingvistike in digitalne humanistike. Je avtorica številnih strokovnih in znanstvenih člankov ter nedavno objavljene monografije o kodnem preklapljanju in odnosu do jezika. Na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani poučuje italijanske jezikovne predmete ter prevajanje med italijanščino in slovenščino. Redno sodeluje z Akademijo za glasbo Univerze v Ljubljani, kjer poučuje italijanščino za pevce. Posveča se prevajanju avdiovizualnih vsebin ter leposlovnih, humanističnih in zgodovinskih besedil, med katerimi velja zlasti omeniti prevod zbirke pisem Giuseppeja Tartinija.