Review report

Manuscript title: "Fully observed shape classification using elastic shape analysis and an application to identifying fossils from the hominin bearing site of Gladysvale, South Africa"

<u>Summary</u>

Thank you for allowing me to review the manuscript "Fully observed shape classification using elastic shape analysis and an application to identifying fossils from the hominin bearing site of Gladysvale, South Africa". This paper is very interesting and contributes greatly to the development of quantitative methods in order to identify fossil specimens. The strength of the paper is the use and the comparison of different methods to generate features of bovid teeth (Elliptical Fourier Analysis and Elastic Shape Analysis, by the use of both size-and-shape and shape only features). On those frameworks, different machine learning algorithms (Random Forest, Support Vector Machines with linear and radial kernel) are then applied to classify the shapes.

I recommend several edits before publication, but ultimately, I recommend this paper to be accepted for publication after considering the comments below.

<u>General suggestions</u>

I should first mention that I have no intention of commenting on the 3. Methods. I am not familiar with the details of those methodologies and have no authority on criticizing them. I will therefore leave this part for the other reviewers and will concentrate on the bovid aspect and the application to the Gladysvale fossil assemblage.

As a general suggestion, I would recommend the authors to change the title by simply replacing the "to identifying" by "to identify". I am not a native English speaker, but I think that it would be correct.

I spotted frequent issues regarding the citation of the tables and figures. You will see below the details in my comments.

The titles of many references lack capital letters, which is problematic for names.

<u>Specific comments</u>

Abstract
Line 23: "A common technique for classifying…" is it that common to use size-and-shape combined with machine learning to identify bovid dental material? I have few examples in mind, but I would not say that it is "common". The real common approach would be the use of comparative osteological collections.

Introduction
Line 50-51: **"**Specifically, animals in the Family Bovidae (antelopes and buffalo) are used because they have strict environmental tendencies such as their habitats, water dependence, migration patterns, etc." I would recommend removing "strict". It sounds a bit weird to me to talk about "strict tendencies", as tendencies are more of a flexible inclination than a definitive rule and because most of bovids are quite adaptative. I would also remove "etc", and finish the sentence like "such as their habitats, water dependence or their migration patterns". Finally, I will add one or two references to support this claim. Like maybe the Kingdon?

Lines 53-54: "Bovids, particularly isolated bovid teeth, are one of the most common fossils recovered at a site in southern Africa" change for "Bovids, particularly isolated bovid teeth, are one of the most common fossils recovered on southern African sites."

Line 56: "(deRuiter et al. [2008], Adams and Conroy [2005], Vrba [1976])." I would add "e.g." before those references, as they are not the only one to use comparative collections.

Lines 57-60: "Recently, geometric morphometrics has been used to help identify isolated bovid teeth by relying on the shape of the outline of the occlusal surface of these teeth, most often through the use of elliptical Fourier analysis (EFA) (Kuhl F [1982], Lestrel [1989], Caple [2017])." There is a problem with the use of references in this sentence. First, the sentence is about the use of geometric morphometrics to specifically identify isolated bovid teeth. However, none of those three references at the end are actually using geometric morphometrics for this purpose. Also, besides Caple (2017), the references are quite old and can't be define as recent work. Second, there no Caple (2017) in the reference list. There is a Byrd, Stephan and Caple (2017) in the reference list, but after a quick check, the order is not correct and should be Caple, Byrd and Stephan (2017), therefore Caple et al. (2017) in the text.

Lines 88-90: "In this work, we derive shape features using elastic shape analysis and use them to classify bovid teeth using modern machine learning approaches." Please remove "modern". The machine learning approaches used in this paper (random forest and support vector) are not really new.

Line 95: "Gladysvale is a hominin bearing site located in the Bloubank Valley of the…" I would recommend to separate the Gladysvale paragraph by making a specific section for it because there is no apparent transition between the previous paragraph and the Gladysvale part, which feels a bit off.

Lines 100-102: "The Bovidae Alcelaphini tribe includes four extant species: *Damaliscus dorcas*, *Alcelaphus buselaphus*, *Connochaetes gnou*, *Connochaetes taurinus*." There are more than four extant species of Alcelaphines. The authors didn't list *Beatragus hunteri* (hirola) or *Damaliscus lunatus* (tsessebe). Moreover, I know that historically, *D. dorcas* is still used in south African palaeontology, but the species is now named *Damaliscus pygargus*. A non-palaeontologist will not know what *D. dorcas* refers to.

Line 102: "The teeth of these four species…" change the four for six.

Lines 104-105: "Thus, many researchers are forced to group them into this ambiguous category" Which ambiguous category are you referring to? The "medium sized alcelaphines"? If it is the case, it could be nice to remind the reader here because there are two sentences between the first mention and the second, so it is difficult to understand what you are talking about.

Lines 113-116: "This study uses the database B.O.V.I.D. (Brophy and Matthews [2022]) which includes extant bovids from 7 tribes and 20 species (Table 1). The reclassifications are used to generate a more precise paleoenvironmental reconstruction." This should belong to the 2. Data section and not the introduction.

Data
Regarding the material, the authors have to make a statement about how they select the teeth specimens regarding their wear stages. Is the sample composed only by adult specimens? If yes, it needs to be stated. It is also necessary to mention if you included any old individual (disappearance of the occlusal cavities) or not. Same for juveniles.

Line 124: "… in Table 1." I believe you meant Table 2, as the table 1 has already been mentioned and refers to the extant bovid species list.

Table 1: change *Damaliscus dorcas* to *Damaliscus pygargus*.

Tribe classification: Shape Features
Line 305: "Figure 8 shows the…" I believe that you meant Figure 5.

Tribe classification: Size-and-Shape Features
Line 331: "Figure 5 shows classification accuracies…" I think you meant Figure 7.

Species classification: Size-and-Shape Features
This section could be more detailed, like the previous ones. For example, in all the previous sections, the specific performance of each method was stated. It would be nice here to specify the performance (%) of RF, OV-PC, I-PC and EFA directly in the text and to better describe the figure 9. Also, there is a typo at the end of the paragraph with an isolated "T".

Application: Reclassification of Gladysvale Alcelaphines
Line 365: "We report classifications at the tribe and level…" there is the "species" missing between "and" and "level".

Line 394: "*Oryx gazealla*" should be written "*Oryx gazella*".

Line 409: change *D. dorcas* to *D. pygargus*.

Line 450: You are referring to the Table 6, which has already been cited while the Table 7 remains uncited in the text. I think you meant Table 7 here.

Lines 468-469: "Accurately and objectively identifying the taxonomic classification of fossil Bovidae teeth is of much import in paleoenvironmental reconstruction." Change to "is of much importance…"?

References
I highlighted in yellow on the PDF file all the name that should have a capital letter.

Line 534: *Connochaetes gnou* should not be in italic if the entire title is in italic.

Lines 596-598: "E.S. Vrba. New fossils of alcelaphini and caprinae (bovidae: Mammalia) from awash, ethiopia, and phylogenetic analysis of alcelaphini. ~~Palaeontologia africana~~. *Palaeontologia africana*, 34:127–198, 1997." Please delete the first Palaeontologia Africana.