In the manuscript entitled "Automated classification of Chandra X-ray point source…" (MN-22-3602-MJ) by Shiva Kumaran et al the authors apply machine learning tools for the automated classification of X-ray point sources, from the Chandra Source Catalogue (CSC 2.0), into various types of astronomical sources like AGN, X-ray emitting stars, young stellar objects etc. The input data includes, in addition to the Chandra X-ray data, multiwavelength data from various surveys. The authors are able to identify a large number of sources at the 3\sigma and 4\sigma level.

The paper is in general well written and the methods followed in the analysis have been described clearly in the earlier part of the paper. The later part needs some changes and improved explanations, which I suggest below. The results are useful, though some further considerations are needed. I should be able to recommend acceptance of the manuscript for publication after the changes are made.

DETAILED COMMENTS:

Section 2.1
The meaning and significance of the variability properties calculated using the three methods needs to be explained.

Section 3.3
In the context of missing values, what is the meaning of imputation? The statement "…we select to impute the missing values using column mode values" is not very meaningful without further explanation, please provide that.

From Figure 1 it is seen that the percentage of missing values is very high in the lower part of the table. Surely, imputation will not work very well in such cases. You would not be able to impute well if the percentage availability were zero. The availability is very small in several of the cases. This matter needs to be discussed well.

In Figure 1, there are only two variability related parameters are included, while three were defined in Section 2.1.

Section 3.4
This Section should only contain the first two paragraphs which describe the upsampling and SMOTE. The other two paragraphs should be transferred to Section 3.5.

The same comments as those for Section 3.3 apply to the class imbalance problem. The authors say that there is a "vast imbalance" in the number of training sources. In such a case, does it make sense to use a technique such as SMOTE to tackle the problem?

This fact needs to be clearly mentioned and discussed. The fourth paragraph of the current Section 3.4 does mention the problem, but it needs elaboration.

Section 3.5

The new Section 3.5 will have the last two paragraphs of the current Section 3.4. The material needs to be rearranged so that the definitions and explanations are all done before the confusion matrix is introduced.

What exactly are the numbers in the confusion matrix in Figure 2? Are these related to precision or recall? How are the numbers calculated? I find that these matters are explained when Figure 5 is described. The explanations should be transferred to the discussion of Figure 2, and need not be repeated later.

In Fig. 2, for the no upsampling case, the numbers for LMXB and later sources are not good. The improvement after using SMOTE does not take the numbers to satisfactory levels. Clearly the data is sufficient to give results at the acceptable level for AGN, Star and YSO and to a lesser extent HMXB, which constitute 10 percent of the training set. The other objects, LMXB, ULX, CV and pulsar together constitute less than 10 percent of the training set, and therefore the results obtained for them are poor. A catalogue of identifications of X-ray point sources made using the ML techniques can be trusted at better than 90 percent for the first three kinds of objects. But the entries for the other objects would have poor confidence and therefore would not be useful.

Section 4

It is said in the fourth paragraph of the section that "…steeper the plot towards unity, higher the predicted CMP and more confident the model. The figure reveals that the LightGBM is the most confident classifier…". But from the figure it is appears the curve for LightGBM is the least steep. Please clarify.

It is seen from Table 7 that use of just the X-ray data already provides precision and recall values close to 90 percent. Use of the multiwavelength data further improves the levels. The results for LMXB and later are not as good. I believe that the results for the four classes with at least 10 percent data fraction (AGN, Star, YSO and HMXB) will improve significantly if the other four kinds of sources are not included in the training and validation. I would like the authors to provide a separate table with just the four top kinds of sources. If the results are good, then a catalogue of such source identified with the X-ray point sources would be very useful.

In the same spirit, the role of various data sets like GAIA, 2MASS etc. in the training needs to be investigated. From Fig 1 it seen that these catalogues have data only for a small percentage of the sources, leading to many missing values. These catalogues may therefore not be adding value to the training, and it could in fact be counterproductive to use them. Training with these data sets omitted needs to be considered.

In Figure 6, some remarks, if possible, on the different shapes of pdf for different class of sources will help.

Section 5

The meaning of the statement "…we will identify those sources which could be assigned to various classes…and thus claim the discovery of many new sources…" is not clear. Are the sources not already identified in the present paper?

General Comment:

There are several minor grammatic errors spread through the manuscript which need to be attend to by the authors and/or the editorial team.