**Reviewers' comments:**  
Reviewer #1: This manuscript requires a complete rewrite and revision to meet the publishing standard of a scientific journal.  
  
The work appears to be a set of R code for data wrangling, i.e. formatting and reorganizing data from other tools, and for data visualization. How generalizable of the code is entirely unclear in the manuscript. The novelty of this work is not clearly described, because many functions discussed in the manuscript are provided by other tools. Is there any novel development of algorithms? If not, is there value in software engineering?  
  
The term "Artificial Intelligence" is in the title, while "Artificial Intelligence" is not seen in the described method.  
  
"Evaluation" in this manuscript appears to be applications to experimental datasets, without describing what is evaluated.  
  
The authors show little academic standard in reporting a computational work. What is the statistical method and its justification in the visualization method? Testing significance in ontology is a mature topic - yet the authors offer no detail on the decision making within the code, and how it's related to prior work.  
  
It is fine to write this as a software paper, even without real novelty in algorithms. The manuscript contains many good ideas, but not developed fully. The authors need to follow scientific conventions and standards to report the work. Visualization does not replace enrichment tests. Data analysis does not replace method evaluation.  
Many confusions in concepts make this manuscript almost incomprehensible. The erroneous use of English is so prevalent that this requires serious editorial help.  
  
  
  
Reviewer #2: The authors describe a new software tool for visualizing metabolomics data, MCnebula. This tool consists of an R package (also called MCnebula), and requires the installation and configuration of some dependent tools, namely Sirius 4.  
  
Here are a few general comments regarding the manuscript:  
(1) Remnants of "track changes" are still visible, e.g. "developed\_" (red underline in abstract). Please remove these.  
  
(2) The manuscript would benefit greatly by a thorough revision specifically focused on the English language. At times, the high amount of English-associated errors makes it difficult to interpret the message of the paper. I have tried, as much as possible, to separate issues associated with English from the scientific contribution described in this paper, and I am confident that English-specific issues in the manuscript can be overcome, provided that the authors make this a priority.  
  
A few general English notes:  
- Avoid using the phrase "in virtue of"  
- Always say "untargeted" or "targeted" in reference to a variety of LC-MS/MS analysis (instead of "untarget" or "target")  
- Avoid using the phrase "in counterpart"  
- Check that all words used are real English words - e.g., "parallelity" is not a word.  
  
(3) In the delivered manuscript packet for review, the figures were not numbered or labeled. It is unclear if this issue is actually due to the submission system (or, perhaps, user error in using the submission system), but this should be corrected.  
  
For this work to be published, I think the tutorial associated with MCnebula has to work properly. I was able to download and install MCnebula, along with the associated Sirius4 dependency. Unfortunately, the tutorial didn't work for me. I treed to follow the instructions provided at http://htmlpreview.github.io/?https://github.com/Cao-lab-zcmu/VIgnette/blob/master/MCnebula\_workflow.html  
I even wrote my own R markdown notebook, trying to adapt the components of this vignette to work on my computer.  
  
To help you in fixing the errors in the tutorial, here is the contents of that notebook:  
---  
---  
title: "MCnebula\_workflow\_tutorial"  
author: "Anonymous\_Reviewer"  
output: html\_document  
---  
  
# Setup  
  
## knitr  
```{r setup, include=FALSE}  
knitr::opts\_chunk$set(echo = TRUE)  
```  
  
## Install MCnebula (if not done already)  
```{r install}  
devtools::install\_github("Cao-lab-zcmu/MCnebula")  
```  
  
## Libraries  
```{r libraries}  
library(MCnebula)  
library(dplyr)  
library(ggplot2)  
library(ggraph)  
library(grid)  
```  
  
## Files and Directories  
```{r directories}  
base\_dir <- getwd()  
mgf.path <- file.path(base\_dir, "raw\_instance", "instance5.mgf")  
  
# version information: sirius-5.5.6-osx64-headless.pkg  
# system information: mac os x Monterey 12.3.1  
sirius\_executable <- "/Applications/sirius.app/Contents/MacOS/sirius"  
  
sirius\_output\_file <- file.path(base\_dir, "test")  
```  
# Workflow  
  
## Run Sirius  
```{r sirius\_prep}  
# data.table::fread(mgf.path, header = F, sep = NULL)  
  
# works  
#cmd <- glue::glue("{sirius} -i {mgf} -o {output} --maxmz 800 formula -c 50 zodiac",  
#cmd <- glue::glue("{sirius} -i {mgf} -o {output} --maxmz 800 formula -c 50 --database canopus",  
  
# does not work  
# cmd <- glue::glue("{sirius} -i {mgf} -o {output} formula -c 10 structure --database pubchem canopus",  
# cmd <- glue::glue("{sirius} -i {mgf} -o {output} --maxmz 800 formula -c 50 zodiac structure canopus",  
# cmd <- glue::glue("{sirius} -i {mgf} -o {output} --maxmz 800 formula -c 50 structure --database canopus",  
  
# given in tutorial  
cmd <- glue::glue("{sirius} -i {mgf} -o {output} --maxmz 800 formula -c 50 zodiac --database canopus",  
  
sirius=sirius\_executable,  
mgf=mgf.path,  
output=sirius\_output\_file)  
  
# print for testing  
#cat(cmd)  
  
system(cmd)  
```  
  
## MC Nebular commands  
```{r mc\_nebular\_commands}  
MCnebula::initialize\_mcnebula(sirius\_output\_file, rm\_mc.set = T)  
  
collate\_structure\_results <- MCnebula::collate\_structure(  
exclude\_element = c("Cl", "S", "P"),  
ppm\_error = 20  
)  
# this errors out. Unable to continue with tutorial.  
  
collate\_pcpp\_data <- MCnebula::collate\_ppcp(  
## due to the size of example dataset, we set much smaller herein  
## min\_possess, 20 or more may be better  
min\_possess = 1,  
## 0.1 may better  
max\_possess\_pct = 0.9  
)  
# this errors out. Unable to continue with tutorial.  
```  
---  
  
  
Reviewer #3:  
Review of manuscript CR-MEDICINE-D-22-00382 entitled "MCnebula: focusing on unknown compound classes by Artificial Intelligence systematic clustering visualization"  
  
Summary  
  
In this manuscript, the authors present a computer tool (MCnebula) for analyzing and visualizing untargeted MS data. This tool allows the characterization of sample unknown compounds by using their fragmentation properties to determine their classes, structures and substructures. The performance of the MCnebula tool is demonstrated using two examples: a human serum metabolomic analysis and a herbal analysis.  
  
Commentaries  
  
In my opinion, the MCnebula tool presented in this manuscript could be useful, and the manuscript describes its operation and performance appropriately. However, I believe this manuscript is more appropriate for a bioinformatics-like journal; therefore, I cannot recommend its publication in CR Medicine.  
  
I have two main concerns regarding the manuscript.  
  
The first one is related to the manuscript text. Unfortunately, I miss a more detailed description of the theoretical basis of the MCnebula algorithm. Now, the MCnebula algorithm seems only a protocol of commands for its use, and I would like to see a more detailed description of its basis.  
  
The second main concern is related to the dependence on the MCnebula software. As described in the text, MCnebula requires inputs from several MS data tools (i.e. Mzmine2, SIRIUS, CSI:finger ID, CANOPUS, …). All these tools are community-supported and probably available for a long time (or alternatives giving the same information). On the one hand, an MCnebula user needs to have solid bioinformatic knowledge and the ability to use several data analyses. On the other hand, I think that this complex workflow will reduce the number of potential users. On the one hand, all the required tools could also be updated, and there will be the uncertainty of whether their outputs will continue to be useful. For instance, MZmine3 version has recently been launched.  
  
Other comments:  
  
I agree with the authors that in recent years there have been huge developments in the field of in silico fragmentation. However, these in silico predictions have been seriously flawed in the past, for example, in the lipids field. Please, discuss.  
Benchmark comparison. The definition of noise to the MS/MS spectra is unclear. Please, give values defining these medium and high levels. Also, describe with examples how were affected these MS/SM spectra. Were these high noise conditions similar to real ones in experimental conditions?  
  
Metabolomic analyses were performed using top intense ions and DDA (human) and top intense (orbitrap plants). If alternative approaches were employed for data acquisition, how would be affected the MCnebula workflow?  
  
All codes used in the manuscript should be provided. Also, check the links for the MASSIVE dataset. Both links are not pointing to the correct dataset.  
  
Check several typos in the manuscript.